Effective Health Care Program

Comparative Effectiveness Review
Number 98

PCA3 Testing for the Diagnosis and Management of Prostate Cancer

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Comparative Effectiveness Review

Number 98

PCA3 Testing for the Diagnosis and Management of Prostate Cancer

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10058-I

Prepared by:
Blue Cross and Blue Shield Association Technology Evaluation Center
Evidence-based Practice Center
Chicago, IL

Investigators:
Linda A. Bradley, Ph.D., FACMG
Glenn Palomaki, Ph.D.
Steven Gutman, M.D., M.B.A.
David J. Samson, M.S.
Naomi Aronson, Ph.D.

AHRQ Publication No. 13-EHC030-EF
April 2013

This report is based on research conducted by the Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10058-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Bradley LA, Palomaki G, Gutman S, Samson DJ, Aronson N. PCA3 Testing for the Diagnosis and Management of Prostate Cancer. Comparative Effectiveness Review No. 98. (Prepared by the Blue Cross and Blue Shield Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-2007-10058-I). AHRQ Publication No. 13-EHC030-EF. Rockville, MD: Agency for Healthcare Research and Quality; April 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an email list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Supriya Janakiraman, M.D., M.P.H.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Claudia J. Bonnell, M.L.S., B.S.N.; Jennifer Brock, M.P.H.; Ryan Chopra, M.P.H.; Michael Douglas, M.S.; Lisa Garofalo, B.A.; Denise Oliansky, M.A., M.L.I.S.; Lisa Sarsany, M.A.; Kathleen M. Ziegler, Pharm.D.

Key Informants

Peter Albertsen, M.D.
Chief, Director of Urology
University of Connecticut Health Center
Farmington, CT

Robert Carey
RPC Associates/Consulting Services
John's Creek, GA

Janet Colli, M.D.
Fellow, Laparoscopic Surgery
Tulane University
New Orleans, LA

John Concato, M.D.
Professor of Medicine (General Medicine)
Yale University
New Haven, CT

William Dotson, Ph.D.
Senior Coordinating Scientist
Public Health Genomics
Centers for Disease Control and Prevention
Atlanta, GA

Peter Gann, M.D., Sc.D.
Professor of Pathology
College of Medicine
University of Illinois at Chicago
Chicago, IL

John L. Gore, M.D., M.S.
Assistant Professor of Urology
University of Washington
Seattle, WA

Roger Klein, M.D., J.D.
Affiliate Associate Professor
Center of Evidence-Based Medicine &
Health Outcomes Research
Clinical and Translational Science Institute
University of South Florida
Morsani College of Medicine
Tampa, FL

Michael LeFevre, M.D., M.S.P.H.
Chief Medical Information Officer
University of Missouri School of Medicine
Columbia, MO

Technical Expert Panel

Peter Albertsen, M.D.
Chief, Director of Urology
University of Connecticut Health Center
Farmington, CT

Todd Alonzo, Ph.D.
Associate Professor of Research
Department of Preventive Medicine
University of Southern California
Arcadia, CA

William Dotson, Ph.D.
Senior Coordinating Scientist
Public Health Genomics
Centers for Disease Control and Prevention
Atlanta, GA

Peter Gann, M.D., Sc.D.
Professor of Pathology
College of Medicine
University of Illinois at Chicago
Chicago, IL

Roger Klein, M.D., J.D.
Affiliate Associate Professor
Center of Evidence-Based Medicine & Health Outcomes Research
Clinical and Translational Science Institute
University of South Florida
Morsani College of Medicine
Tampa, FL

Stephen Spann, M.D., M.B.A.
Professor and Chair
Family/Community Medicine
Baylor College of Medicine
Houston, TX

Thomas Trikalinos, M.D., Ph.D.
Professor of Medicine
Co-Director, Tufts Evidence-based Practice Center
Institute for Clinical Research and Health Policy Studies
Boston, MA

Peer Reviewers

Daniel Barocas, M.D., M.P.H.
Assistant Professor Medicine
Vanderbilt University
Nashville, TN

Carl Gjertson, M.D.
Assistant Professor, Surgery
University of Connecticut
Farmington, CT

Andrzej Kosinski, Ph.D.
Associate Professor
Biostatistics/Bioinformatics
Duke University
Durham, NC

Timothy O'Leary, M.D., Ph.D.
Director, Clinical Science Research and Development
Veterans Affairs
Washington, DC

Estelle Russek-Cohen, Ph.D.
Deputy Director
Division of Biostatistics
U.S. Food and Drug Administration
Rockville, MD

PCA3 Testing for the Diagnosis and Management of Prostate Cancer

Structured Abstract

Objectives. We performed a comparative effectiveness review that examined the use of the prostate cancer antigen 3 (PCA3) gene in improving initial or repeat biopsy decisions in patients identified at risk for prostate cancer, or in improving decisionmaking about treatment choices (e.g., active surveillance vs. aggressive therapy) in patients with prostate cancer positive biopsies. Comparators included total prostate specific antigen (PSA) elevations, free PSA, PSA density, PSA velocity, externally validated nomograms, complexed PSA, and multivariate models.

Data sources. We searched PubMed® and Embase® from January 1, 1990, to August 8 and August 15, 2011, respectively, and updated through May 15, 2012. We searched the Cochrane Database of Systematic Reviews with no date restriction and updated. A grey literature search included databases with regulatory information, clinical trial registries, abstracts/conference papers, grants and federally funded research, and manufacturer information.

Review methods. Inclusion criteria required PCA3 and at least one comparator to be measured in the same cohort in one of the three clinical settings: at-risk men considering initial biopsy; at-risk men considering repeat biopsy; and men with prostate cancer making treatment decisions based on risk categorization. Data were extracted by one reviewer and audited by a second. Analyses were matched by comparing within study differences between PCA3 and a comparator. Modeling was used to smooth consensus ROC curves and to address issues relating to verification bias. Diagnostic accuracy studies were assessed for quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. Strengths of evidence were judged high, moderate, low, or insufficient according to Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria and the AHRQ "ethods Guide for Medical Test Reviews."

Results. After exclusion of six studies with strong likelihood of containing duplicate data, 24 studies provided data that could be used to address diagnostic accuracy (Key Questions [KQ] 1 and 2); 13 studies addressed decisionmaking based on risk stratification criteria (KQ 3). All studies were of poor quality. Comparison of PCA3 to total PSA (tPSA) had the most available studies (22) but was subject to spectrum, verification, and sampling biases; the latter two were addressed in the analyses. We observed that: (1) PCA3 is more discriminatory for detecting cancers (i.e., at any sensitivity, the specificity is higher, or at any specificity, the sensitivity is higher) than tPSA elevations; (2) this finding appears to apply to both initial and repeat biopsies; and (3) PCA3 and tPSA are relatively independent predictors. However, strength of evidence was low. For all other diagnostic accuracy comparisons, and all intermediate and long-term health outcomes, the strength of evidence was insufficient. For treatment decisionmaking in men with positive biopsy, in all comparisons for intermediate and long-term health outcomes, the strength of evidence was found to be insufficient.

Conclusions. For diagnostic accuracy, there was a low strength of evidence that PCA3 had better diagnostic accuracy for positive biopsy results than tPSA elevations, but insufficient evidence that this led to improved intermediate or long-term health outcomes. For all other settings, comparators, and outcomes, there was insufficient evidence.

Contents

Tables

Figures

Appendixes

Executive Summary

Background

Cancer of the prostate is the second most common cancer and the second leading cause of cancer deaths in men in the United States.[1,2] Most patients have slow-growing tumors, and may live for years with no or minimal effects, ultimately dying of other causes.[3,4] The lifetime risk of being diagnosed with prostate cancer is 16 percent, but lifetime risk of dying from the disease is only 3 percent.[1,3,5] However, some patients have aggressive tumors that spread beyond the prostate, resulting in significant morbidity and death. A challenge in managing clinically localized disease is distinguishing between men who have aggressive disease and need immediate therapy, and those who have less aggressive disease that can be safely managed by active surveillance.

Production of serum total prostate specific antigen (tPSA) was found to be increased in men with prostate cancer as many as 5 to 10 years prior to symptoms of clinical disease.[6] The rationale for initiating prostate cancer screening using tPSA was to reduce the prevalence of advanced prostate cancer and prostate cancer-related mortality through early detection, and improve quality of life.[3,7,8] Prostate cancer mortality has decreased,[1,2] but at what cost in overdiagnosis and potential harms related to treatment?[11] Also, issues such as who to test, when to test and retest, and the most effective clinical tPSA threshold continue to be debated. A recent U.S. Preventive Services Task Force recommendation concluded that the potential benefits do not outweigh the harms.[9] However, the balance of benefits and harms of tPSA screening remains controversial.[9,10]

In 1999, researchers reported that the prostate cancer antigen 3 gene (PCA3; also known as DD3), was highly overexpressed in prostate cancer relative to normal prostate or benign prostatic hyperplasia tissue.[12] Subsequently, PCA3 tests on messenger RNA from urine were developed.[13,14] Two proposed intended uses of PCA3 and comparator tests were to inform decisionmaking about initial or repeat biopsy of men with elevated tPSA and/or other risk factors. The third was to inform decisions about management and treatment (e.g., active surveillance, prostatectomy, radiotherapy) by classifying disease in men with positive biopsies as insignificant or aggressive.

The U.S. Food and Drug Administration (FDA) recently approved a PCA3 assay for use in men 50 years of age or older who have had one or more previous negative biopsies, but did not have a finding of atypical small acinar proliferation in the most recent biopsy. The intended use of the test is to inform decisionmaking about repeat biopsy.

Scope and Key Questions

Biomarker comparators for detection of prostate cancer at biopsy considered in this review are tPSA, specific isoforms of tPSA, and validated risk-assessment calculators or nomograms.

- Serum tPSA is widely available as a screening and monitoring test using set (e.g., 2.5 or 4 ng/mL) or age-specific cutoffs.[15]
- One tPSA isoform is free PSA, reported as a ratio of free to total PSA or percent free PSA (%fPSA). Low levels (less than 25 percent) are associated with cancer and high levels with benign disease. Percent fPSA may be useful in decisionmaking about biopsy, particularly for men whose tPSA levels are in the “grey zone” (2.5 to 10 ng/mL).

- A second isoform is PSA bound to serum antiproteases, or complexed PSA (cPSA). Data are limited but performance may be similar to %fPSA.
- PSA density is the ratio of tPSA concentration to prostate volume. Addition to tPSA may improve the prediction of positive biopsy or insignificant cancer, but this has not been confirmed.
- PSA velocity and doubling time are measures of longitudinal increases in tPSA. Utility of PSA velocity for predicting positive biopsy or insignificant cancer is not clear.[17] PSA doubling time has value for monitoring patients with advanced or recurrent cancer.[18]
- Externally validated nomograms are risk assessment tools that combine multiple clinical and laboratory risk factors to inform clinical decisionmaking about biopsy, risk classification, and/or treatment options. Despite variability and lack of validation in some cases, such tools may provide better information than use of individual markers.

For risk classification, PCA3 comparators in a prognostic workup include Gleason score, prostate volume, risk factors, biochemical markers, and clinical/pathological staging.

The Key Questions (KQs) relate to the three proposed scenarios described above:

KQ1: In patients with elevated PSA and/or an abnormal digital rectal examination who are candidates for initial prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decision making about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life, and potential harms?

KQ 2: In patients with elevated PSA and/or an abnormal digital rectal examination who are candidates for repeat prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decisionmaking about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life, and potential harms?

KQ 3: In patients with a positive biopsy for prostate cancer who are being evaluated to distinguish between insignificant/indolent and aggressive disease, what is the effectiveness of using PCA3 testing alone, or in combination with the standard prognostic workup (e.g., tumor volume, Gleason score, clinical staging) or monitoring tests (e.g., PSA, PSA velocity), with regard to diagnostic accuracy (clinical validity) for aggressive (high-risk) prostate cancer, intermediate outcomes (e.g., improved decisionmaking about prognosis and triage for active surveillance and/or aggressive treatment), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life, and potential harms?

Corresponding analytic frameworks are presented in the full report.

Methods

Literature Search Strategy

We searched PubMed®, Embase®, and the Cochrane Central Register of Controlled Trials for the timeframe January 1, 1990 to August 15, 2011; updated searches were performed for the timeframe ending May 15, 2012. The grey literature searches included regulatory information, clinical trial registries, conference papers, and selected Web sites. We included studies that were in English, reported primary data, addressed KQs, and fulfilled the criteria for: (1) study design (matched studies in the same clinical setting in which PCA3 and comparators were assessed in

all men in a study population); (2) study subjects/populations (at-risk men or men with a positive biopsy); (3) study interventions (biomarker testing, biopsy, risk classification); (4) study comparators; and (5) intermediate (diagnostic accuracy, impact on decisionmaking, harms) and long-term (e.g., mortality, morbidity, function, quality of life, harms) outcomes.

For title/abstract and full-article review, one reviewer read and determined eligibility and a second reviewer audited a subset of abstracts (and all marked uncertain) and all full articles; discrepancies were resolved by discussion or a third reviewer when needed. Data were extracted by a single reviewer, and then fully audited by a second senior reviewer. Disagreements were resolved through review team discussion.

Quality Assessment of Individual Studies and Strength of Evidence

In adherence with the Methods Guide,[19] grading the methodological quality of individual comparative studies was performed based on study design-specific criteria. The quality of diagnostic accuracy studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool.[20] The QUADAS ratings were summarized into general quality classes of good, fair, and poor.[19] The strength of evidence for outcomes was evaluated using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system.[21] GRADE addresses four domains of evidence (risk of bias, consistency, directness, and precision), and rates each body of evidence as High, Moderate, Low, or Insufficient.[19,21] In all cases, two independent reviewers assessed the quality of individual studies and the strength of evidence. Discordant decisions were resolved through discussion or third-party adjudication.

Data Synthesis

For KQ 1 and KQ 2, PCA3 scores were evaluated against all comparators for which published data were available. Analyses included clinical sensitivity, clinical specificity (or the false positive rate equal to 1-specificity), and positive and negative predictive values. When data were available, the following analyses were performed for PCA3 and one or more comparators:

- Differences in area under the receiver operating characteristic (ROC) curve (AUC)[a], including direction and magnitude of differences.
- Reported medians and standard deviations in positive and negative biopsy populations (reported as z scores), including direction and strength of effect.
- Performance at a PCA3 cutoff score of 35 (sensitivity and specificity).
- Receiver operator characteristic (ROC) curves (sensitivity and specificity), to evaluate fixed specificities and compare corresponding sensitivities.
- Regression analysis using regression coefficients and associated relative odds ratios.

Based on the limited number of studies identified that address KQ 3, we anticipated focusing on a qualitative analysis (e.g., descriptive narrative, summary tables, identification of themes in content).

[a]Area under the curve (AUC) is a common metric that measures the accuracy of diagnostic tests, that is the ability of the tests to discriminate those who have (or will develop) the outcome of interest from those who do not have (or will not develop) the outcome. An AUC of 1.0 indicates a perfect test, and an AUC of 0.5 indicates a worthless test.

Applicability

Applicability of the results presented in this review was assessed in a systematic manner using the PICOTS framework (Population, Intervention, Comparison, Outcome, Timing, Setting).

Results

Detailed description of analyses with tables and figures are included in the full report.

Results of Literature Searches

Our inclusion criteria restricted the analyses to matched studies that provide data on PCA3 and at least one other comparator in the same patient population. Population matching was preserved by computing differences between PCA3 test results and comparator test results within each study, and comparing these differences across studies. Searches identified 1,556 citations, of which 220 underwent full text review and 42 were included. Grey literature search identified 1 additional study for a total of 43.

Potential Biases in Included Studies

Subjects in the included studies were drawn from academic medical centers where patients with elevated tPSA results and/or other risk factors were seeking referral or specialty care. Observational studies of such opportunistic cohorts are subject to specific biases.

- **Verification bias:** Men are most often offered prostate biopsy based on the extent of tPSA elevation. Higher tPSA levels indicate higher likelihood of prostate cancer, and men are more likely to undergo prostate biopsy if tPSA is high (e.g., 10-20 ng/mL) rather than closer to cutoffs (e.g., 3-4 ng/mL). Estimates of sensitivity and specificity at select tPSA cutoffs will be impacted in studies in which biopsy decisions are tPSA-related. If those not accepting biopsy are considered missing, this is considered "partial verification" bias.
- **Spectrum bias:** Convenience samples can also predispose to spectrum effects, as they may represent men at higher risk of prostate cancer than the total cohort of screened men. Of more concern is that the range of severity of disease predicted by PCA3 and comparators could be different, for example, if men positive on one test were found to have different characteristics or severity of disease than those positive on another test.
- **Sampling bias:** A subset of studies restricted enrollment to tPSA results in the "grey zone" (e.g., 2.5 ng/mL to < 10 ng/mL). The effect was to reduce both the prevalence of disease in the study group and tPSA test performance, as those men with higher tPSA levels (where tPSA is most predictive) are not enrolled in the study. This bias cannot be avoided by statistical analysis, but was addressed by stratifying analyses and summarizing "grey zone" studies separately.

KQ 1: Initial Biopsy

Two matched studies reported results in populations where all men were having initial biopsies.[22,23] Both reported comparisons of PCA3 with tPSA and %fPSA; one[22] also reported on PSA density. All studies were graded as poor, and strength of evidence was rated insufficient

because there were too few data for reliable interpretation. No studies addressed other comparators or outcomes; all other comparisons were graded insufficient.

KQ 2: Repeat Biopsy

Seven matched studies addressed diagnostic accuracy for KQ 2, reporting results in populations where all men were having a repeat biopsy.[24-30] Five studies[24-28] reported on PCA3 and tPSA, four[25-27,29] on %fPSA, one on PSA velocity,[26] and two on externally validated nomograms.[24,30] However, the numbers of comparisons possible for each of these matched analyses remained small. For example, one of three tPSA studies providing AUC data restricted recruitment to men with tPSA levels in the "grey zone."[27] No studies addressed other comparators or outcomes. Strength of evidence was insufficient for all comparisons.

KQ 1 and 2: Initial and Repeat Biopsies

In addition to the 9 studies described above, another 15 studies provided matched PCA3 and tPSA data and reported the proportion of men having initial and repeat biopsies.[31-45] Given inadequate strength of evidence for analyses focused on men with initial or repeat biopsy only, we examined all studies to determine suitability for a combined analysis (Table A). Using the most commonly reported comparator and analysis (tPSA and AUC), we performed a regression analysis of AUC difference (PCA3 – tPSA) versus the proportion of study subjects on whom prostate biopsy. Based on linear regression, the slope was not significant (p=0.97), indicating no significant relationship between biopsy status and AUC difference for PCA3 versus tPSA elevations. Three of the 15 studies also reported AUCs stratified by initial and repeat biopsy status that could replace the composite AUCs; analysis with these data showed that the slope was again not significant (p=0.81).

Fourteen studies also provided ROC curves for both PCA3 and tPSA. Regression analysis of (PCA3 – tPSA) sensitivities at a specificity of 50 percent versus biopsy status again showed little or no association between biopsy history and relative performance of PCA3 and tPSA (p=0.79). No similar analyses can be made for any other comparator for diagnostic accuracy. This was considered sufficient to proceed with a combined analysis for KQ 1/KQ 2, without the biopsy history restriction. The same regression analysis conducted in different datasets (e.g., including/ excluding "grey zone" studies, stratified by assay type) consistently found no significant slope. In addition, very similar median AUC differences (PCA3 – tPSA) were found for studies enrolling all men having initial biopsy and studies enrolling all men having repeat biopsy.

Total PSA (tPSA) Elevations, PCA3 Score, and Diagnostic Accuracy for Combined KQ 1/KQ 2 Analysis

Subsets of 20 studies provided sufficient data to compare the diagnostic accuracy of PCA3 with tPSA elevations, using the five described analyses (Table A). We identified two important biases: verification bias and sampling bias. Verification bias occurred for the comparator tPSA (and related measures), as the extent of those elevations was often the basis for deciding on biopsy. Modeling was used to account for the potential impact of verification bias. A sampling bias occurred for the comparator tPSA (and related measures) when some studies only enrolled men with tPSA elevations in the "grey zone" (e.g., upper limit of 10 ng/mL). This results in diminished test performance for tPSA, as it is most predictive of a positive biopsy when very elevated. This bias was accounted for by stratification.

Figure A shows the consensus observed ROC curves for PCA3 and tPSA using data from the 13 studies with suitable analyses. Based on other internal analyses and modeling, it was possible to generate smooth overlapping logarithmic Gaussian curves that fitted these observed data well. Table B shows select data from this modeling that compares the ability of PCA3 and tPSA to identify prostate cancer among men at increased risk. The first row of Table B shows that at a set false-positive rate of 80 percent (specificity of 20 percent), the corresponding sensitivity for PCA3 scores is 95.8 percent. This is 5.1 percentage points higher than the 90.7 percent sensitivity for tPSA measurements. The table then compares sensitivities of the two markers at lower false positive rates. The bottom of Table B displays differences in the false positive rates for selected sensitivities. This is just another way to view the data from the fitted ROC curves.

Both Figure A and Table B indicate that PCA3 is associated with higher sensitivity at any given specificity than tPSA elevations, and higher specificity at any given sensitivity. This combined approach made it possible to reliably compare PCA3 and tPSA measurements for diagnostic accuracy. Quality of all individual studies was poor and strength of evidence was low (Table A). For all other comparators (Table A), and all other intermediate (impact on decisionmaking, harms of biopsy) and long-term outcomes (morbidity/mortality, quality of life, potential harms), analyses were not possible or were constrained by variability of study populations and limited numbers of studies. All individual studies were graded poor and strengths of evidence for all other outcomes were insufficient.

KQ 3: Testing PCA3 and Comparators To Identify Men With Insignificant Cancer Who May Benefit From Active Surveillance

Thirteen studies were identified that addressed KQ 3 and reported on PCA3 and other preoperative/pretreatment markers for stratification of prostate cancer by risk.[22,42,43,46-55] Two studies based analyses on biopsy markers without prostatectomy[22,42] and eight reported prostatectomy results as an endpoint. Two studies[46,47] were conducted on subjects in an active surveillance program and included short-term followup. One[46] predicted outcome based on identification of micrometastases through measurement of tPSA and PCA3 in lymph node extracts, and reported decreased 4- to 6-year biochemical recurrence-free survival in patients with identified micrometastases. Another[47] reported 2-year followup of progression from active surveillance to treatment in men with prostate cancer, based on results of yearly biopsy.

Quality of all studies was poor. Strength of evidence was insufficient for diagnostic accuracy due to the inability to compare or combine the two studies on different sample types, using different parameters (e.g., sensitivity/specificity, mean/median biomarker levels) and addressing different outcomes. For this outcome: risk of bias was high; consistency was unknown with two studies; directness was indirect; and precision could not be assessed (imprecise). Strength of evidence was insufficient for any other outcomes or comparators, as no studies were identified.

Key Findings and Strength of Evidence

Strength of evidence was insufficient for KQ 3 and for all comparators and outcomes for KQ 1 and KQ 2 except the comparison of PCA3 and tPSA for the outcome of diagnostic accuracy (Figure A, Table A, Table B). Among men at risk, PCA3 was more discriminatory for detecting prostate cancer at biopsy than tPSA elevations. The finding that the relative performance of PCA3 versus tPSA elevations does not appear to be dependent on biopsy history is a new

observation that could impact future studies. The quality of all studies was poor. The strength of evidence was considered low.

Table A. PCA3 versus comparators—analyses and strength of evidence for the intermediate outcome of diagnostic accuracy

Comparators	tPSA[a]	%fPSA[a]	PSAD[a]	EVN[a]	Multivariate Models Including tPSA[a]	cPSA[a]	tPSA DT and tPSA Velocity[a]
GRADE: Risk of Bias	High	High	High	High	High	---	---
GRADE: Consistency	Consistent, with 24 studies	Inconsistent[b], with 7 studies	Unknown[b], with 3 studies	Unknown[b], with 4 studies	Unknown[b], with 3 studies	---	---
GRADE: Directness	Indirect	Indirect	Indirect	Indirect	Indirect	---	---
GRADE: Precision	Precise	Imprecise	Imprecise	Imprecise	Imprecise	---	---
GRADE: Dose-Response Relationship	Present	---	---	---	---	---	---
GRADE: Strength of Association	Weak	---	---	---	---	---	---
Strength of Evidence (GRADE)[c]	Low	Insufficient	Insufficient	Insufficient	Insufficient	Insufficient	Insufficient
KQ 1 and KQ 2 Area Under the Curve	n=20[22,24-28,31-40,42-45]	n=5[23,25-27,32]	n=3[22,28,36]	n=3[24,30,37]	0	0	0
Reported Mean/SD	n=8[23,24,26,32,35-37,43]	n=4[22,26,32,37]	n=2[22,36]	0	0	0	0
Performance at a PCA3 cutoff of 35	n=9[22,25,28,31-34,36,41]	n=1[22]	n=2[22,38]	n=1[37]	0	0	0
ROC Curves–Sensitivity/ Specificity	n=14[22,24,25-28,31-38,40]	n=4[23,25,27,32]	n=3[22,28,36]	n=2[36,37]	0	0	0
Regression Analysis	n=2[22,37]	n=3[22,25,37]	0	n=2[24,30]	n=3[22,25,37]	0	0

%fPSA = percent free PSA; cPSA = complexed PSA; EVN = externally validated nomograms; tPSA = total prostate specific antigen; PSAD = PSA density; PSAV = tPSA velocity or doubling time; ROC = receiver operating characteristic; SD = standard deviation; tPSA = total prostate specific antigen

[a]Corresponds to KQ 1 and KQ 2.

[b]Consistency could not be assessed due to insufficient data from comparable studies, or because studies did not report results in a consistent manner.

[c]GRADE assessment of strength of evidence for each outcome for each comparator is based on assessment of the evidence for four domains: risk of bias; consistency of effect size/direction, directness of the evidence-health outcome link; and precision (degree of certainty) of effect estimates (e.g., estimates of sensitivity, AUC differences). Based on the domains, GRADE strength of evidence categories are Insufficient, Low, Moderate and High.

Table B. Comparison of PCA3 and tPSA measurements to identify men with prostate cancer, holding constant either the false-positive rate (1-specificity) or sensitivity

Part 1: False-Positive Rate (FPR) Held Constant			
FPR (1-specificity) %	**PCA3 Scores Sensitivity %**	**tPSA Elevations Sensitivity %**	**Effect With PCA3: % Improvement in PCa Detection**
80	95.8	90.7	5.1
70	92.0	84.3	7.7
60	87.2	77.6	9.6
50	81.1	68.8	14.0
40	73.7	59.7	14.0
30	63.8	48.4	15.4
20	51.6	36.4	15.2
Part 2: Sensitivity Held Constant			
Sensitivity %	**PCA3 Scores FPR (1- Specificity) %**	**tPSA Elevations FPR (1-Specificity) %**	**Effect With PCA3: % Reduction in Biopsies**
95	77.7	88.2	10.5
90	65.6	78.2	12.6
85	56.3	71.5	15.2
80	48.5	63.6	15.1
70	36.2	51.2	15.0
60	26.6	40.3	13.7
50	18.9	31.2	12.3

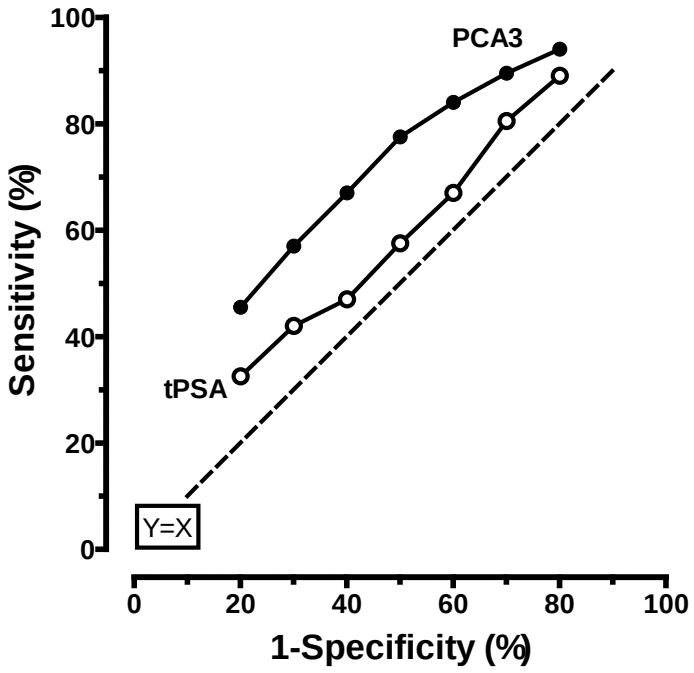
DR = proportion of biopsy positive men with a PCA3 score or tPSA elevation at or above the cutoff level; FPR = proportion of biopsy negative men with a PCA3 score or tPSA elevation at or above the cutoff level; tPSA = total prostate specific antigen

Discussion

An important consideration in this conclusion was the potential for spectrum bias, and the associated indirectness of evidence for identifying positive biopsy status. We made the underlying assumption that not all positive biopsies are equal. For example, identifying a positive biopsy associated with a high Gleason score or specific pathological findings may be considered to be clinically more valuable than one with only a low Gleason score. None of the included studies provided a two-way cross tabulation of PCA3 and tPSA positive and negative test results among biopsy positive patients. Of most interest would be the clinical finding for the cases in the off-diagonal (when one test is positive and the other negative).

For KQ 3, the literature review revealed few relevant matched studies and a lack of clinical followup after patients were placed into risk categories defined by the results of PCA3 and other biomarker and pathological tests. In 11 of 13 studies, a reference clinical endpoint (or validated surrogate) was lacking. The quality of all individual studies was poor and strength of evidence was insufficient. It is likely that more time will be needed for studies to assess the diagnostic accuracy of predicting long-term outcomes for patients based on categorization as having low-risk or high-risk disease.

Figure A. Observed consensus receiver operator characteristic (ROC) curves for PCA3 scores and tPSA elevations



DR = proportion of biopsy positive men with a PCA3 or tPSA value above the cutoff level; FPR = false positive rate or 1- specificity (proportion of biopsy negative men with a PCA3 or tPSA value above the cutoff level); PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen

Note: The open circles (solid line) indicate the consensus observed performance of PCA3 scores, while the filled circles (solid line) indicated the matched consensus observed tPSA performance. The dashed line indicated where the sensitivity equals 1-specificity, indicating a test with no predictive ability. For each study, the sensitivities of PCA3 and tPSA at preselected false positive (1-specificity) rates (x-axis) were estimated from the published ROC curves; median consensus sensitivities were derived for each (1-specificity) rate (y-axis).

Applicability

The populations studied in the included articles were largely drawn from academic medical centers where patients with elevated tPSA results and/or other risk factors (e.g., positive digital rectal examination, family history, race) often seek, or are referred for, specialty care. Performance of PCA3 and comparators in a broader range of health care settings may differ from that described in this review. It is not yet clear how PCA3, alone or in combination with other biomarkers/risk factors would be integrated into diagnostic or management pathways. The level of acceptance by physicians (and consumers) may well be impacted by Food and Drug Administration approval of a test kit to inform decisions about repeat biopsy in men with a specific clinical history that includes previous negative biopsies. While there was evidence that PCA3 performed better than tPSA with regard to diagnostic accuracy as a secondary test for men with increased risk, it is important to note that neither PCA3 nor tPSA have high performance. A combination of biomarkers and other risk information may be needed to improve overall performance in predicting prostate cancer at biopsy, or informing treatment based on risk classification. The intermediate outcome of diagnostic accuracy is key, as improvement could directly impact the number of biopsies performed in men without prostate cancer and the number of men with prostate cancer who are missed. It is also important to understand other potential

harms, as well as the impact of the information on decisionmaking. The effect of even a great test is limited if uptake is low. Longer-term outcomes are challenging, due to the difficulty of following patients and collecting the necessary information.

Research Gaps

With the exception of analyses that include PCA3 and tPSA for the intermediate outcome of diagnostic accuracy, evidence was insufficient to answer the KQs. These questions, therefore, articulate remaining gaps in evidence. Other gaps in knowledge include:

- How much improvement in diagnostic accuracy is needed for any new test to impact biopsy decisionmaking?
- What is the potential of adding PCA3 alone or with other biomarkers to change decisionmaking in practice?
- How does PCA3 compare with the two more frequently used add-on tests (free PSA, PSA velocity) that have appeared in guidance documents?
- Matched studies not derived from "convenience" populations (e.g., biopsy referral centers), and more data on how key demographic factors (family history, race) impact on the performance of PCA3 and comparators.
- Outcome studies to determine how well PCA3 and other comparators used to categorize risk as insignificant/indolent or aggressive to predict the behavior of tumors over time.
- A range of methodological and statistical questions relating to modeling, assessing impact of verification bias, identifying most effective cutoffs for tests based on ROC analysis, and designs for future studies.

Conclusions

For diagnostic accuracy, there was a low strength of evidence that PCA3 had better diagnostic accuracy than tPSA elevations, but insufficient evidence that this led to improved intermediate or long-term health outcomes. In men at risk for prostate cancer based on elevated serum tPSA levels and/or suspicious digital rectal exam or other risk factor (e.g., family history), PCA3 was found to be more discriminatory for predicting prostate cancer at biopsy than tPSA elevations (i.e., at any sensitivity, the specificity is higher, or at any specificity, the sensitivity is higher). The finding that the relative performance of PCA3 versus tPSA elevations is not dependent on biopsy history (i.e., initial biopsy or repeat biopsy after one or more negative biopsies) is a new observation that allowed more studies on KQ 1 and KQ 2 to be combined for analyses. Strength of evidence was insufficient for all other comparators and all other outcomes of interest in KQ 1 and KQ 2.

Eleven of 13 studies addressing KQ 3 lacked a defined reference clinical endpoint (or validated surrogate), and the other two addressed different outcomes. Strength of evidence was Insufficient. There was insufficient evidence for all other comparators and for all other outcomes of interest in KQ 3. With one exception, these three questions continue to identify important gaps in knowledge, with other gaps identified in the review. Current uncertainty about the utility of tPSA screening for prostate cancer[30-34] makes understanding followup tests (e.g., PCA3, other biomarkers, and algorithms) for assessing risk prior to biopsy and/or treatment particularly important.[35,36]

References

1. Brawley OW. Prostate cancer epidemiology in the United States. World J Urol. 2012 Apr;30(2):195-200. PMID: 22476558.
2. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. CA Cancer J Clin. 2010 Sep-Oct;60(5):277-300. PMID: 20610543.
3. National Comprehensive Cancer Network Guidelines—Prostate Cancer Early Detection, Version I.2011. www.NCCN.org.
4. Freedland SJ. Screening, risk assessment, and the approach to therapy in patients with prostate cancer. Cancer. 2011 Mar 15;117(6):1123-35. PMID: 20960523.
5. Yin M, Bastacky S, Chandran U, et al. Prevalence of incidental prostate cancer in the general population: a study of healthy organ donors. J Urol. 2008 Mar;179(3):892-5; discussion 95. PMID: 18207193.
6. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003 95(12):868-78. PMID: 12813170.
7. Basch E, Oliver TK, Vickers A, et al. Screening for prostate cancer with prostate-specific antigen testing: American Society of Clinical Oncology Provisional Clinical Opinion. J Clin Oncol. 2012 Jul 16; PMID: 22802323.
8. Catalona WJ, Smith DS, Ratliff TL, et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med. 1991 Apr 25;324(17):1156-61. PMID: 1707140.
9. Moyer VA, on behalf of the USPSTF. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2012 Jul 17;157(2):120-34. PMID: 22801674.
10. Schroder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med. 2012 Mar 15;366(11):981-90. PMID: 22417251.
11. Chou R, Croswell JM, Dana T, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2011 Dec 6;155(11):762-71. PMID: 21984740.
12. Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. 1999 Dec 1;59(23):5975-9. PMID: 10606244.
13. Durand X, Moutereau S, Xylinas E, et al. Progensa PCA3 test for prostate cancer. Expert Rev Mol Diagn. 2011 Mar;11(2):137-44. PMID: 21405964.
14. Makarov DV, Loeb S, Getzenberg RH, et al. Biomarkers for prostate cancer. Annual review of medicine. 2009 Feb;60:139-51. PMID: 18947298.
15. Oesterling JE, Jacobsen SJ, Chute CG, et al. Serum prostate-specific antigen in a community-based population of healthy men. Establishment of age-specific reference ranges. JAMA. 1993 Aug 18;270(7):860-4. PMID: 7688054.
16. Lee R, Localio AR, Armstrong K, et al. A meta-analysis of the performance characteristics of the free prostate-specific antigen test. Urology. 2006 Apr;67(4):762-8. PMID: 16600352.
17. Vickers AJ, Savage C, O'Brien MF, et al. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. J Clin Oncol. 2009 Jan 20;27(3):398-403. PMID: 19064972.
18. Vickers AJ, Brewster SF. PSA velocity and doubling time in diagnosis and prognosis of prostate cancer. Br J Med Surg Urol. 2012 Jul 1;5(4):162-68. PMID: 22712027.
19. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; August 2011. AHRQ Publication No. 10(11)-EHC063-EF. www.effectivehealthcare.ahrq.gov

20. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006 6(Mar 6):529-36. PMID: 16519814.

21. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--Agency for Healthcare Research and Quality and the Effective Health-care Program. J Clin Epidemiol. 2010 May;63(5):513-23. PMID: 19595577.

22. de la Taille A, Irani J, Graefen M, et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. J Urol. 2011 Jun;185(6):2119-25. PMID: 21496856.

23. Ferro M, Bruzzese D, Perdona S, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta. 2012 Aug 16;413(15-16):1274-8. PMID: 22542564.

24. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008 Oct;180(4):1303-8; discussion 08. PMID: 18707724.

25. Aubin SM, Reid J, Sarno MJ, et al. PCA3 molecular urine test for predicting repeat prostate biopsy outcome in populations at risk: validation in the placebo arm of the dutasteride REDUCE trial. J Urol. 2010 Nov;184(5):1947-52. PMID: 20850153.

26. Auprich M, Augustin H, Budaus L, et al. A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. BJU Int. 2012 Jun;109(11):1627-35. PMID: 21939492.

27. Ploussard G, Haese A, Van Poppel H, et al. The prostate cancer gene 3 (PCA3) urine test in men with previous negative biopsies: does free-to-total prostate-specific antigen ratio influence the performance of the PCA3 score in predicting positive biopsies? BJU Int. 2010 Oct;106(8):1143-7. PMID: 20230386.

28. Wu AK, Reese AC, Cooperberg MR, et al. Utility of PCA3 in patients undergoing repeat biopsy for prostate cancer. Prostate Cancer Prostatic Dis. 2012 Mar;15(1):100-5. PMID: 22042252.

29. Pepe PAragona F. PCA3 score vs PSA free/total accuracy in prostate cancer diagnosis at repeat saturation biopsy. Anticancer Res. 2011 Dec;31(12):4445-9. PMID: 22199313.

30. U.S. Food and Drug Administration. Summary of Safety and Effectiveness Data: PROGENSA PCA3 Assay, 2012. www.accessdata.fda.gov/cdrh_docs/pdf10/P100033b.pdf.

31. Adam A, Engelbrecht MJ, Bornman MS, et al. The role of the PCA3 assay in predicting prostate biopsy outcome in a South African setting. BJU Intl. Epub 2011 Apr 20; PMID: 21507188.

32. Bollito E, De Luca S, Cicilano M, et al. Prostate cancer gene 3 urine assay cutoff in diagnosis of prostate cancer: a validation study on an Italian patient population undergoing first and repeat biopsy. Anal Quant Cytol Histol. 2012 Apr;34(2):96-104. PMID: 22611765.

33. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol. 2008 Apr;179(4):1587-92. PMID: 18295257.

34. Goode RR, Marshall SJ, Duff M, et al. Use of PCA3 in detecting prostate cancer in initial and repeat prostate biopsy patients. Prostate. 2012 May 14; PMID: 22585386.

35. Nyberg M, Ulmert D, Lindgren A, et al. PCA3 as a diagnostic marker for prostate cancer: a validation study on a Swedish patient population. Scand J Urol Nephrol. 2010 Dec;44(6):378-83. PMID: 20961267.

36. Ochiai A, Okihara K, Kamoi K, et al. Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. Int J Urol. 2011 Mar;18(3):200-5. PMID: 21332814.

37. Perdona S, Cavadas V, Di Lorenzo G, et al. Prostate cancer detection in the "grey area" of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. European Urol. 2011 Jan;59(1):81-7. PMID: 20947244.

38. Rigau M, Morote J, Mir MC, et al. PSGR and PCA3 as biomarkers for the detection of prostate cancer in urine. Prostate. 2010 Dec 1;70(16):1760-7. PMID: 20672322.

39. Roobol MJ, Schroder FH, van Leeuwen P, et al. Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: exploring the value of PCA3 for a first-line diagnostic test. European urology. 2010 Oct;58(6):475-81. PMID: 20637539.

40. Schilling D, Hennenlotter J, Munz M, et al. Interpretation of the prostate cancer gene 3 in reference to the individual clinical background: implications for daily practice. Urol Int. 2010 85(2):159-65. PMID: 20424427.

41. Wang Y, Sun G, Pan JG, et al. Performance of tPSA and f/tPSA for prostate cancer in Chinese. A systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2006 9(4):374-8. PMID: 16926855.

42. Cao DL, Ye DW, Zhang HL, et al. A multiplex model of combining gene-based, protein-based, and metabolite-based with positive and negative markers in urine for the early diagnosis of prostate cancer. Prostate. 2011 May 15;71(7):700-10. PMID: 20957673.

43. Hessels D, van Gils MP, van Hooij O, et al. Predictive value of PCA3 in urinary sediments in determining clinico-pathological characteristics of prostate cancer. Prostate. 2010 Jan 1;70(1):10-6. PMID: 19708043.

44. Mearini E, Antognelli C, Del Buono C, et al. The combination of urine DD3(PCA3) mRNA and PSA mRNA as molecular markers of prostate cancer. Biomarkers. 2009 Jun;14(4):235-43. PMID: 19489685.

45. Ouyang B, Bracken B, Burke B, et al. A duplex quantitative polymerase chain reaction assay based on quantification of alpha-methylacyl-CoA racemase transcripts and prostate cancer antigen 3 in urine sediments improved diagnostic accuracy for prostate cancer. J Urol. 2009 Jun;181(6):2508-13; discussion 13-4. PMID: 19371911.

46. Kusuda Y, Miyake H, Kurahashi T, et al. Assessment of optimal target genes for detecting micrometastases in pelvic lymph nodes in patients with prostate cancer undergoing radical prostatectomy by real-time reverse transcriptase-polymerase chain reaction. Urol Oncol. 2011 May 18; PMID: 21600799.

47. Tosoian J, Loeb S. PSA and beyond: the past, present, and future of investigative biomarkers for prostate cancer. The Scientific World Journal. 2010 10(1919-31. PMID: 20890581.

48. Auprich M, Chun FK, Ward JF, et al. Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. Euro Urol. 2011 Jan;59(1):96-105. PMID: 20980098.

49. Durand X, Xylinas E, Radulescu C, et al. The value of urinary prostate cancer gene 3 (PCA3) scores in predicting pathological features at radical prostatectomy. BJU Int. 2012 Jul;110(1):43-49. PMID: 22221521.

50. Liss MA, Santos R, Osann K, et al. PCA3 molecular urine assay for prostate cancer: association with pathologic features and impact of collection protocols. World J Urol. 2011 Oct;29(5):683-8. PMID: 21152924.

51. Ploussard G, Durand X, Xylinas E, et al. Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance. European urology. 2011 Mar;59(3):422-9. PMID: 21156337.

52. van Poppel H, Haese A, Graefen M, et al. The relationship between Prostate Cancer gene 3 (PCA3) and prostate cancer significance. BJU Int. 2012 Feb;109(3):360-6. PMID: 21883822.

53. Vlaeminck-Guillem V, Devonec M, Colombel M, et al. Urinary PCA3 score predicts prostate cancer multifocality. J Urol. 2011 Apr;185(4):1234-9. PMID: 21334023.

54. Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. J Urol. 2008 Nov;180(5):1975-8; discussion 78-9. PMID: 18801539.

55. Nakanishi H, Groskopf J, Fritsche HA, et al. PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. J Urol. 2008 May;179(5):1804-9; discussion 09-10. PMID: 18353398.

Introduction

Background

Burden of Illness

Prostate cancer is the most commonly diagnosed non-skin cancer and the second leading cause of cancer death in the United States.[1,2] In 2010, there were 217,730 new cases of prostate cancer in the U.S. and about 32,000 prostate cancer-related deaths.[1] More than 60 percent of cases occur in men 70 years of age or older.[2,3] Established risk factors for prostate cancer are increasing age, family history and race/ethnicity.[2-4] Family history in an affected brother, father or multiple family members, significantly increases the risk of developing prostate cancer.[2,5] Black/African American men are 1.6 times more likely to be diagnosed with prostate cancer and have a 2.5 times higher risk of prostate cancer-related death.[2,5] Asian American and Native American men have significantly lower risk.[2,5] Other risk factors (e.g., dietary factors, increased hormone levels, familial disposition) have been proposed but not clearly established.[2-4,6]

The disease is unpredictable, with the rate of tumor growth varying from very slow to moderately fast.[5,7] Most patients have slow-growing tumors, and may live for years with no or minimal effects, ultimately dying of other causes.[5,7,8] Although the lifetime risk of being diagnosed with prostate cancer is 16 percent, the lifetime risk of dying from the disease is only 3 percent.[2,5,9] Even more striking is the prevalence of occult disease. About one-third of men older than age 60, and half of men older than age 70, were found to have prostate cancer at autopsy.[5,10]

However, some patients have aggressive tumors that spread beyond the prostate, resulting in significant morbidity and death. Therefore, the key diagnostic challenge in dealing with prostate cancer is deciding which patients to biopsy and when. The most pressing challenge in managing clinically localized disease is distinguishing between men who have aggressive disease and need aggressive therapy and men who have less aggressive disease and can be safely managed by active surveillance.

Screening for Prostate Cancer

Production of serum total PSA (tPSA) was found to be increased in men with prostate cancer, with elevation of tPSA preceding clinical disease by as much as 5 to 10 years.[11] Screening programs that used the tPSA test have been in place since the early 1990s, with the rationale of reducing the prevalence of advanced prostate cancer and prostate cancer-related mortality through early detection, and improving quality of life.[5,12,13] Over time, concerns have been raised about both false-positive tPSA screening results (overdiagnosis) and false-negative results (missed diagnosis). Overdiagnosis has been defined as detection by screening of disease that would not have become symptomatic or clinically significant.[14] Men with false-positive screening results may undergo one or more biopsies with negative results, raising the potential for adverse events.

It has been estimated that as many as 75 percent of eligible men in the U.S. have undergone at least one tPSA test.[15] However, the balance of benefits and harms of tPSA screening remains controversial. Prostate cancer-related mortality has decreased by about 40 percent after screening implementation (1991 to 2007).[1,2,16,17] It is not known how much this drop in mortality can be attributed to screening, and how much part aggressive and improved treatment, or to other factors such as increased risk of death being attributed to other chronic illnesses (e.g., heart

disease).[2,18-20] Wider availability of screening has also resulted in a notable rise in prostate cancer incidence, and in men being diagnosed at an earlier age and cancer stage (i.e., localized disease).[2,4,10] False positives related to the poor specificity of the test, and treatment of increased numbers of men with early stage disease, have contributed to concerns about the health benefits of tPSA screening. Based on an updated systematic review of the available evidence[19], the U.S. Preventive Services Taskforce recently released a draft recommendation against tPSA screening in asymptomatic men in the general U.S. population.[21] They concluded with "…moderate certainty that the harms of PSA-based screening for prostate cancer outweigh the benefits."[21]

Several professional associations guidelines have supported tPSA screening in asymptomatic men age 50 years of age or older, and in younger men in higher-risk populations.[5,12,22,23] Potential harms have been acknowledged, and "routine" screening not necessarily proposed, but it has been recommended that men be provided with information on the potential risks and benefits of tPSA screening, and then allowed to make a decision about screening in consultation with their physician.

Diagnosis and Management of Prostate Cancer

Clinical action points for tPSA screening results have ranged from 2.5 to 10 ng/mL, with 4 ng/mL commonly used as defined in the initial FDA approval for this intended use. Criteria for immediate consideration of biopsy in screen positive men have varied, but have almost always included men with tPSA results greater than10 ng/mL, and may include men with tPSA results between four and 10 ng/mL, with or without suspicious digital rectal examination findings.[5,24] Decisions about biopsy in patients with elevated (greater than 4 ng/mL) or intermediate (also referred to as "grey zone") tPSA results (e.g., 2.5 to 4 ng/mL) may also be impacted by other considerations, such as age, family history, race and results of followup testing with tPSA and/or other biomarkers (e.g., %fPSA, PSA density, PSA velocity) and/or use of individualized risk assessment tools or nomograms (e.g., PCPT Risk Calculator[25]).[7,15,26] However, the most effective approaches for use of biomarkers and risk assessment tools/nomograms remain controversial, and evidence of impact on decisionmaking and subsequent short- and long-term improvement in clinical outcomes (e.g., function, morbidity, mortality) are lacking. Decisions about biopsy may also be impacted by other risk factors, comorbidities, or patient and physician preferences.

Biopsy

Performance of needle biopsy leads to pathologic examination of tissue cores (minimally 6 cores, 12 are recommended) to identify the presence or absence of cancer, the percent of the tissue core that is cancer and the Gleason score (an assessment of tissue differentiation).[3,24,27] In addition to anxiety and discomfort, identified risks of biopsy have included infection (with increase in cases of antimicrobial resistance), fever, rectal bleeding, hematuria, vasovagal episodes, hematospermia and dysuria.[5,28]

Classification systems have been developed to designate biopsy-positive men as high risk or low risk relative to the likelihood of disease progression without treatment. Such risk classification can inform decisions about management, such as treatment versus active surveillance,[29,30] and multiple treatment options (e.g., surgery vs. radiation).[5,27] There is considerable interest in the identification of new biomarkers, or effective combinations of biomarkers and/or other risk factors, to better inform prebiopsy decisionmaking.[31]

Treatment Options

For patients with insignificant disease (i.e., low grade, low volume tumors with specific pathologic characteristics, one option is active surveillance.[14,20,27,32] Instead of working to eradicate the tumor, the patient defers treatment and begins ongoing surveillance that minimally includes serial tPSA and PSA velocity testing and repeated biopsies. If there is evidence of conversion to more aggressive disease (i.e., increasing tPSA levels or PSA velocity, upgrading in tumor stage or volume), surgery and other options can be considered.

For patients with "high-risk prostate cancer" of varying risk strata (i.e., high grade, high volume, multifocal tumor or other characteristics believed to be associated with aggressive disease), or those unwilling to accept the risk associated with active surveillance, there are well-established treatment options.[7,27] As examples:

- **Prostatectomy** involves complete excision of the prostate. Prostate tissue from radical prostatectomy undergoes clinical staging and gross and histopathological examination.[3,24,27,33] Quantitative tools for risk stratification are available to estimate the likelihood of prostate cancer recurrence post-prostatectomy or other treatment.[25,34-36] Complications of prostatectomy include urinary incontinence and erectile dysfunction. The 10-year rate of prostate cancer-related death ranged from one to eight percent, depending on classification of the tumor as high or low risk.[37]
- **Interstitial brachytherapy** is a choice for patients with low to moderate risk disease, and involves implanting high-dose radioactive seeds into the prostate. Adverse effects include lower urinary tract symptoms, obstructive or irritative prostatitis, and later onset of erectile dysfunction. Ten-year disease free survival is reported to be about 85 percent.[38]
- **External-beam radiotherapy** is commonly used in combination with androgen therapy to treat high-risk disease. Adverse effects include irritative voiding symptoms, hemorrhagic cystitis or proctitis, and risk of a second malignancy of the bladder and rectum. Reported 10-year disease-free survival is about 88 percent.[37]
- **Other techniques** for whole prostate treatment include cryotherapy and high-intensity focused ultrasound (HIFU).

While biomarkers and risk assessment tools/nomograms are being used to identify men with high- and low-risk cancers, evidence of their effectiveness based on long-term clinical outcomes is needed.

Development of a New Biomarker: PCA3

In 1999, researchers reported identification of the differential display 3 gene (*DD3*), highly overexpressed in prostate cancer tissue but having little or no expression in normal prostate tissue or benign prostatic hyperplasia tissue.[39] Subsequently renamed the prostate cancer antigen 3 gene (PCA3), PCA3 is a noncoding mRNA mapped to chromosome 9q21-22.[40,41] Since 2002, quantitative methods to measure PCA3 mRNA in urine samples have been developed and improved[40,42] (see Table 4 in Results). Early investigators found that prostate manipulation led to a general release of mRNA. Therefore, expression of another prostate-specific gene needed to be quantified to correct for the number of prostate cells present in the urine. One study[43] had investigated the expression of the gene that encodes PSA, *KLK3*, in prostate tissue. They reported that PSA mRNA expression had been shown to be relatively constant in normal prostate cells, with only a weak (~1.5-fold) down-regulation of PSA expression in prostate cancer cells. They further noted the other studies that had shown that levels of protein and mRNA expression were not necessarily parallel. Therefore, PSA mRNA was chosen as the "housekeeping" gene

against which PCA3 mRNA results were normalized.[40,44-46] In most assays, the ratio of PCA3 mRNA copies per mL and PSA mRNA copies per mL is multiplied by 1,000 to provide a PCA3 "score."[40,44-46]

Stability of mRNA has been shown to be a source of variability for some tests, but use of a detergent based stabilization buffer[47] and shipping frozen samples have improved sample quality. Sokoll et al.[48] reported the first multicenter study of PCA3 analytical performance in 2008 using the Gen-Probe assay and concluded that the assay performs well and is insensitive to preanalytical factors.

On February 17, 2012, Gen-Probe reported that they had received U.S. Food and Drug Administration (FDA) approval for the PROGENSA® PCA3 assay. The FDA specified that the test's intended use was for men 50 years of age or older who had one or more previous negative biopsies (and no finding of atypical small acinar proliferation in the most recent biopsy) and are being considered for a repeat biopsy. A negative PROGENSA PCA3 assay is noted to be associated with decreased likelihood of a positive biopsy. However, the label specifies that "...the performance of the assay has not been established in men for whom a repeat biopsy was not already recommended." PCA3 testing is also offered by reference laboratories as laboratory-developed tests (i.e., tests developed by and used at a single laboratory testing site). With FDA approval, it is reasonable to anticipate that the PCA3 test could become more widely available throughout the U.S.

In summary, the upregulation of PCA3 mRNA expression in prostate cancer tissue provided a rationale for detecting a small number of cancer cells within the background of a large number of normal or benign prostatic hypertrophy cells.[40,44] Three potential intended uses for the PCA3 test have been proposed: 1) to inform decisions about when to biopsy patients at-risk and when to wait; 2) to inform decisions about when to rebiopsy patients at-risk and when to wait (the claim currently approved by FDA for the PROGENSA® PCA test); and 3) to determine, in patients with positive biopsies, whether the disease is insignificant/indolent or aggressive, so that an optimal treatment plan can be developed.

Selected PCA3 Comparators

Total PSA (tPSA)

In 1989, Catalona et al.[13] initiated a multicenter population-based study examining the use of tPSA and digital rectal exam (DRE) for prostate cancer screening. Based on the findings of enhanced early prostate cancer detection, the FDA approved a PSA assay for prostate cancer screening in 1994, and defined the upper limit of normal as 4 ng/mL.[41] Numerous studies have followed, and the test is widely available as a screening and monitoring test. A subsequent meta-analysis of studies on PSA as a screening test showed that a significant number of prostate cancers would be missed using 4 ng/mL as a cutoff.[49] As a result, some laboratories subsequently offered testing at lower cutoffs (e.g., 2.5 or 3.0 ng/mL). Lower cutoffs increased the yield of positive cancers, but also increased the false-positive rate and lead to negative biopsies and the potential for subsequent followup with associated risk of clinical harms.[8,19] Others have used age-specific tPSA cutoffs to improve sensitivity in younger men and improve specificity in older men (e.g., cutoffs of 2.5, 3.5, 4.5, and 6.5 for men 40-49, 50-59, 60-69, and 70-79 years of age, respectively). [5,50] However, NCCN Prostate Cancer Early Detection guidelines note that investigations of tPSA age-specific ranges have resulted in "equivocal results,"[5] and clinical utility remains uncertain.

In spite of its importance in health care, serum tPSA is not a standardized analyte, meaning that there is not a national requirement for quality specifications defined by either the Clinical Laboratory Improvement Act (CLIA) regulations or by the FDA. Assay-specific and site-specific differences in serum tPSA test performance in proficiency testing have been reported for more than 15 years.[51-53] In 1999, Klee et al. reported that plus or minus 10 percent bias ranges for PSA correspond to -19.9 percent to +20.4 percent variation in patient classification.[54] Redesign of proficiency testing (PT) materials was reported to improve College of American Pathology survey outcomes,[51] and these specimens are likely to be the best available for monitoring testing in U.S. laboratories.[55] However, a 2006 review of PT data for tPSA found that available methods could be assumed correct less than 40 percent of the time at 6.5 ng/mL, and only 30 to 40 percent of the time at 19 ng/mL.[55] They concluded that test results were insufficiently reliable when applying uniform national cutoffs, and could cause many false-positive results.

In summary, for KQ 1 and KQ 2, nearly all study subjects have tPSA levels of at least 2 to 4 ng/mL or higher (i.e., screen positive). Therefore, the actual comparator for PCA3 is not a positive or negative tPSA but, among those with positive results, the extent of the tPSA elevation. For example, risk of a prostate cancer being identified among men with levels between 4 and 6 ng/mL is lower than among men with tPSA levels greater than 10 ng/mL.

Free PSA (%fPSA)

Serum PSA circulates in blood as different isoforms, mainly as unbound or free PSA (fPSA) and a form that complexes with serum antiproteases (complexed PSA or cPSA).[37] The ratio of fPSA to tPSA levels is reported as percent fPSA. High levels of %fPSA are associated with benign prostatic disease, while low levels are associated with cancer.[56] Percent fPSA increases with age and prostate volume, and decreases as total PSA increases.[41] In addition, %fPSA is less stable than tPSA, and requires processing with 24 hours of collection.

One meta-analysis of %fPSA studies[57] reported that, in the tPSA diagnostic grey zone of 2 to 10 ng/mL, addition of %fPSA testing can reduce the number of negative biopsies while maintaining a high cancer detection rate, but a second meta-analysis reported that %fPSA is a useful addition to tPSA testing in only one part of the grey zone.[58] However, FDA has approved fPSA testing for men with tPSA levels in the 4-10 ng/mL range. The NCCN Prostate Cancer Early Detection guidelines includes %fPSA in its diagnostic algorithm for early detection of prostate cancer, as part of decisionmaking about biopsy or rebiopsy.[5] They suggest that patients with "grey zone" tPSA levels and %fPSA of 10 percent or less are candidates for biopsy or repeat biopsy; patients with %fPSA greater than 25 percent be should be followed closely (e.g., DRE, tPSA, %fPSA, PSA velocity), and that patients with intermediate values be informed of choices.[5]

PSA Density

Benson et al.[59] described the concept of PSA density (PSAD) as the ratio of tPSA concentration to prostate volume, but requires transrectal ultrasound or magnetic resonance imaging to assess prostate volume. They identified differences in mean PSAD between men with prostate cancer and men with BPH. Subsequent studies suggested a modest improvement in diagnostic accuracy when PSAD was added to tPSA values, but more recent studies failed to confirm the value of PSAD. Consequently, its current clinical use appears limited.[41,60] One recent report has suggested this measurement could be used to predict clinical pathological features of disease.[61]

PSA Velocity and Doubling Time

PSA velocity and PSA doubling time are measures of longitudinal increases in tPSA. PSA velocity is defined as the rate of change of tPSA levels in a specified period, typically reported as ng/mL per year.[41,62,63] PSA doubling time is defined as the time it takes (e.g., months) for the tPSA level to increase by a factor of two. Both can pose challenges in practice, as they have been defined in many ways, with variability in number of tPSA measures needed to calculate a dynamics metric, the time between measures, and the statistical method for estimating change.[62,63,64]

Guidelines for early detection of prostate recommend that PSA velocity be considered in both consideration of biopsy and followup.[5] However, systematic reviews of pretreatment use of PSA velocity found evidence that this biomarker is not useful in informing decisionmaking about biopsy, eligibility for active surveillance or post-treatment prognosis.[62-64] Studies have shown that PSA doubling time has prognostic value for metastasis-free and prostate cancer-related survival in monitoring of patients with advanced or recurrent prostate cancer.[64]

Complexed PSA

Another PSA isoform considered in early detection of prostate cancer is measurement of the PSA bound to serum antiproteases, termed complexed PSA (cPSA). A recent systematic review[57] concluded that cPSA and %fPSA showed equivalent effects. The authors cautioned that a lack of detail on study methodology and the relatively small number of studies warranted caution in the interpretation of the findings.

Externally Validated Nomograms

Current decisionmaking about risk of prostate cancer and whether to biopsy or rebiopsy has not been standardized, but depends on consideration of a variety of clinical factors (e.g., age, family history, race) and laboratory test results.[26] Recently, attention has been directed at development of risk algorithms, nomograms or artificial neural networks that combine multiple clinical and laboratory risk factors to create a cumulative risk score that informs clinical decisionmaking.[15,26,34,65] These risk assessment tools are intended to exploit the incremental value of running multiple tests, each with independent contributions to the estimation of the risks of biopsy or treatment outcomes for patients. Risk factors often included were tPSA level, age, race and family history, and other biomarkers that vary by study. A systematic review concluded that these tools produced improvements in area under the receiver operator characteristics (ROC) curve (AUC), when compared with tPSA levels alone, but noted that many nomograms were not externally validated. When external validation was taken into account, benefits of nomogram use were decreased, although still statistically significant.[15] Despite variation in development and validation, a recent systematic review suggested that these tools tend to provide more accurate diagnostic predictions for cancer-positive biopsies than the use of tPSA testing or other factors alone.

Criteria for Distinguishing Insignificant Prostate Cancer

Three criteria have been commonly recognized to identify candidates for active surveillance: low-grade disease, low-volume disease, and low tPSA levels. However, eligibility criteria vary from study to study and site to site.[20] The most established criteria for identifying patients with "insignificant" disease are those proposed by Epstein.[16,30,66] Epstein includes prostate biopsy criteria (clinical stage T1c; PSA density <0.15 ng/mL; no biopsy Gleason pattern 4 or 5; fewer

than three positive cores; and <50 percent cancer per core), as well as criteria predicting pathologic characteristics of radical prostatectomy tissue (organ-confined disease; tumor volume less than 0.5 cm^3; Gleason score of 6 or less without Gleason pattern 4 or 5).[14,67] Original and modified Epstein criteria are common inclusion measures for studies of active surveillance.[20]

Comparison of criteria for identifying low- to high-risk tumors has been complicated by the development of multiple predictive and prognostic nomograms/risk assessment tools, with little or no comparison and standardization of underlying assumptions. Commonly used tools include the D'Amico risk classification, the Kattan nomogram, the Partin tables, the Prostate Cancer Prevention Trial risk calculator and the Nakanishi and Chun nomograms.[14,15,25,66,68] Understanding the relative performance of measurements to identify candidates for active surveillance or treatment is complicated by the use of varying criteria to identify risk progression, including changes in tPSA levels, histological grade, extent of biopsy core involvement, and/or clinical stage.[20] Longitudinal studies are needed to determine long-term clinical outcomes of patients classified as low and high risk, including those who immediately entered treatment or initially chose active surveillance (possibly progressing to treatment).

Key Questions In patients with elevated PSA and/or an This comparative effectiveness review addresses three Key Questions (KQs). The first two relate to the use of the urine PCA3 test and other biomarker tests to predict detection of prostate tumor at biopsy or rebiopsy of men at risk based primarily on elevated tPSA and/or suspicious DRE. A recognized problem in tPSA screening-based diagnosis of prostate cancer is the high rate of false-positive results that can lead to a relatively high number of negative biopsies. The third KQ concerns the use of the urine PCA3 test and other biomarker tests and pathological markers to classify the patient as low or high risk. This review will not address the merits or limitations of prostate cancer screening.

The KQs relate to three proposed scenarios in which this testing may be used:

KQ 1. In patients with elevated PSA and/or an abnormal DRE who are candidates for initial prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decision making about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life and potential harms?

KQ 2. In patients with elevated PSA and/or an abnormal DRE who are candidates for repeat prostate biopsy (when all previous biopsies were negative), what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decisionmaking about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life and potential harms?

KQ 3. In patients with a positive biopsy for prostate cancer who are being evaluated to distinguish between indolent and aggressive disease, what is the effectiveness of using PCA3 testing alone, or in combination with the standard prognostic workup (e.g., tumor volume, Gleason score, clinical staging) or monitoring tests (e.g., PSA, PSA velocity), with regard to diagnostic accuracy (clinical validity) for aggressive (high risk) prostate cancer, intermediate outcomes (e.g., improved decisionmaking about prognosis and triage for active surveillance and/or aggressive treatment) and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life, and potential harms?

The proposed KQs were posted for public comment on the Effective Health Care Program Web site (www.effectivehealthcare.ahrq.gov) from May 4, 2011, to June 1, 2011. A total of eight

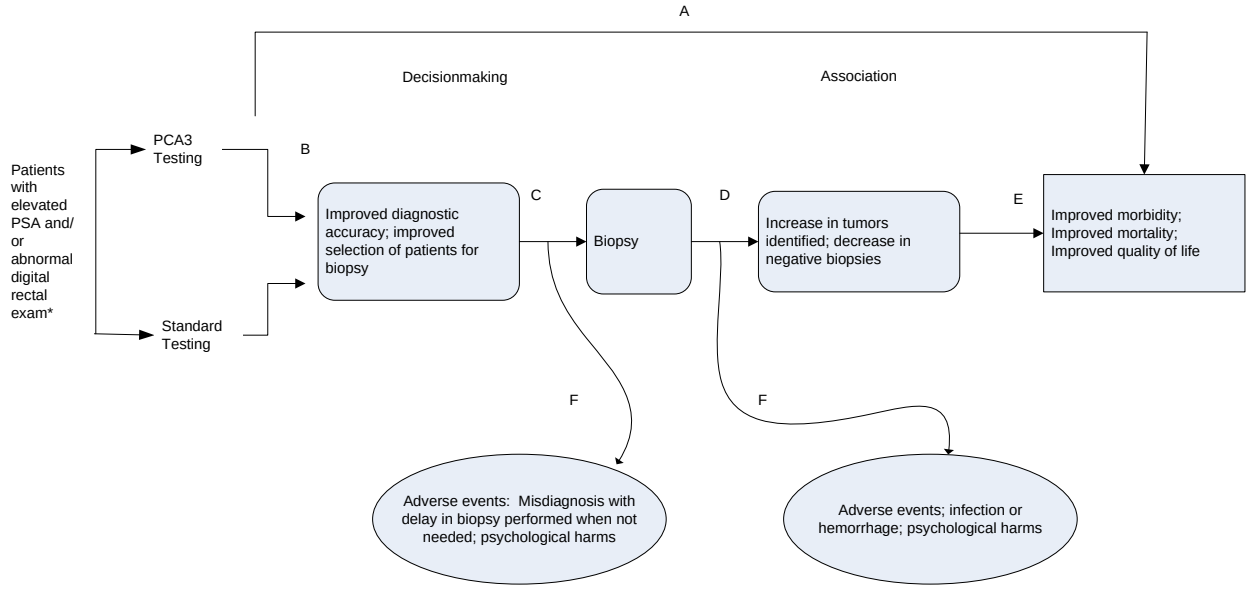
comments were received. No respondent suggested a specific change in the questions, although several noted that data concerning the use of PCA3 testing were currently most compelling for decisionmaking about repeat biopsy in patients screened with a PSA test and a DRE. At least two comments were directed at the likelihood that our review would not be able to address the long-term outcomes of interest (e.g., mortality, morbidity, quality of life). One commentator addressed the value of PCA3 test results in multispecialty team decisionmaking and noted that this test should be evaluated in patients receiving treatment to aid decisions about management changes. Based on the public comments received, no changes were made to the KQs.

Analytic Frameworks

Two analytical frameworks (AFs) were developed for this review, one for KQ 1 and KQ 2, and a second for KQ 3. The first AF (Figure 1) addresses KQ 1 and KQ 2, and applies to men who are at risk for prostate cancer based on elevated tPSA results and/or abnormal DRE, and are having either an initial (KQ 1) or repeat (KQ 2) biopsy. The AF in Figure 1 depicts the comparative effectiveness of using PCA3 testing (alone or in combination with other biomarkers) and other standard tests (e.g., tPSA, %fPSA) to predict intermediate and long-term health outcomes of interest. In Figure 1, direct evidence of the impact of testing on health outcomes (e.g., mortality/ morbidity, quality of life) is shown by Link A. In the indirect chain of evidence, Link B addresses an intermediate outcome, the diagnostic accuracy (clinical validity) of the PCA3 and its designated comparators in predicting positive biopsies. Link C addresses the impact of test results on the decision to proceed to an initial prostate biopsy, which, in turn, impacts other intermediate outcomes (Link D) and leads to long-term health outcomes (Link E) that determine the utility of the tests. Link F on the left addresses potential harms related to the effect testing has on the biopsy decision; Link F on the right focuses on clinical (e.g., bleeding, infection) and psychosocial (e.g., anxiety, quality of life) harms related to the biopsy procedure and/or repeated biopsy.

The second analytic framework (Figure 2) addresses KQ 3, and applies to men who have had a biopsy result that is positive for prostate cancer. Figure 2 depicts the comparative effectiveness of using PCA3 testing (alone or in combination with other biomarkers) and other commonly used tests (e.g., Gleason score and other pathological markers, percent positive cores) on intermediate and long-term health outcomes of interest. Direct evidence of the impact of testing on health outcomes (e.g., mortality, morbidity, function, quality of life) is shown by Link A. In the indirect chain of evidence, Link B addresses the diagnostic accuracy (clinical validity) of the test's categorization of tumors as aggressive or insignificant/indolent in predicting health outcomes. Link C addresses the impact of test results on decisionmaking related to prognosis and triage for active surveillance versus aggressive treatment. This link, in turn, impacts other intermediate outcomes (Link D) and leads to the long-term health outcomes (Link E) that determine both the clinical validity of tests in accurately categorizing risk and the utility of treatment based on risk categorization. Link F on the left addresses potential harms related to the effect testing has on treatment decisions (e.g., incorrect therapy); Link F on the right focuses on clinical (e.g., incontinence, impotence, bowel dysfunction) and psychosocial (e.g., anxiety, self-image, quality of life) harms related to management and treatment.

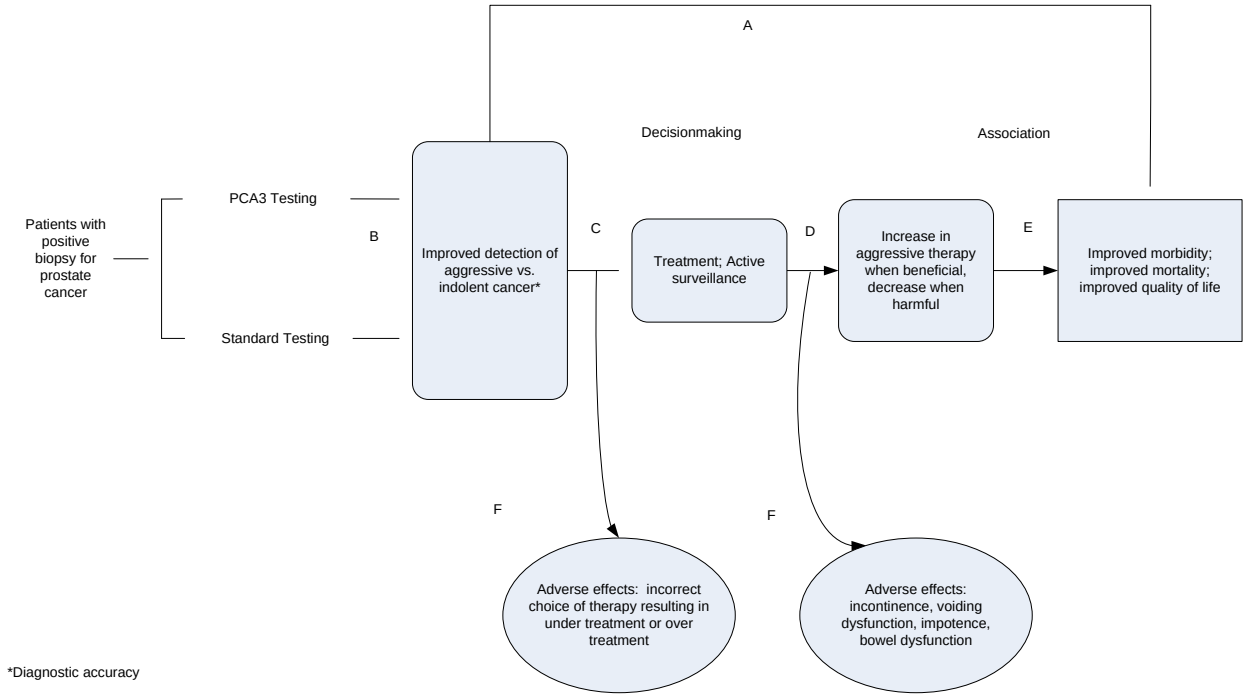
Figure 1. Analytic framework for PCA3 as a diagnostic indicator for biopsy or rebiopsy in patients with elevated PSA and/or abnormal digital rectal examination (KQ 1 and KQ 2)



*Patients may be evaluated for initial biopsy or rebiopsy after one or more negatives.

PSA = prostate specific antigen; PCA3 = prostate cancer antigen 3 gene

Figure 2. Analytic framework for PCA3 used to distinguish insignificant/indolent versus aggressive prostate cancer (KQ 3)



*Diagnostic accuracy

PCA3 = prostate cancer antigen 3 gene

PICOTS Framework

The PICOTS (Population, Intervention, Comparator, Outcome, Timing, and Setting) for the KQs follow:

Population(s)

KQ 1: Adult male patients who are candidates for initial prostate biopsy based on elevated prostate-specific antigen (tPSA) and/or abnormal digital rectal examination (DRE).

KQ 2: Adult male patients with one or more previous negative prostate biopsies who are candidates for repeat biopsy based on elevated tPSA and/or abnormal DRE.

KQ 3: Adult male patients with a positive prostate biopsy.

Interventions

- Testing for the prostate cancer antigen 3 gene (*PCA3*) mRNA alone or in conjunction with comparator tests
- Prostate biopsy
- Prostatectomy

Comparators

- KQs 1 and 2
  - Total PSA
  - Percent free PSA
  - PSA velocity or doubling time
  - PSA density
  - Complexed PSA
  - Externally validated nomograms
- KQ 3
  - Total PSA
  - Percent free PSA
  - PSA velocity or doubling time
  - Externally validated nomograms
  - Gleason score
  - Stage
  - Prostate volume
  - Epstein and other risk criteria
  - Other pathological markers

Outcomes

KQs 1 and 2

- *Long-term health outcomes*: Prostate cancer-related mortality, morbidity, function, quality of life (measured with validated instruments), and harms related to PCA3 testing and subsequent interventions (e.g., biopsy, surveillance, treatment).
- *Intermediate outcomes*: Diagnostic accuracy; impact on decisionmaking that leads to reduction in the number of negative biopsies and increase in the identification of prostate tumors.

KQ 3

- *Long-term outcomes*: Prostate cancer-related mortality, morbidity, function, quality of life (measured with validated instruments) and harms related to PCA3 testing and subsequent interventions (e.g., repeat biopsy, active surveillance, and treatment).
- *Intermediate outcomes*: Diagnostic accuracy; impact on decisionmaking that provides information on prognosis and informs treatment decisions.

Timing

- Any duration of followup will be evaluated.
- Timing of studies related to successive generations of PCA3 and PSA assays will be considered as part of quality assessment and as a potential source of heterogeneity.

Setting

All settings.

Scope of the Review

Despite the large body of published literature on prostate cancer screening with tPSA, the value of early intervention remains controversial.[19,21,69] However, the burden of prostate cancer and the efforts to effectively diagnose and treat the disease are substantial. Consequently, there is continuing interest in identifying and validating biomarkers that can improve the clinical specificity and sensitivity of the prostate cancer diagnostic pathway (i.e., predicting prostate cancer in at risk men) or prognosis (i.e., classifying prostate tumors as high or low risk, informing treatment decisions). Introduction of such a biomarker, alone or in combination with other biomarkers or risk factors, has the potential to reduce the current uncertainty in decisionmaking, which may lead to improved health outcomes with lower risk of associated harms. However, systematic review of the evidence supporting the diagnostic accuracy and utility of each intended use is needed to ensure an overall net balance of benefits over harms, and reduce the risk of introducing new unanticipated harms.

This is a comparative effectiveness review of testing with prostate cancer antigen 3 (PCA3), alone or in combination with other markers (comparators), in three proposed intended uses. The KQs addressed how PCA3 testing effectiveness compares with other markers: (1) alone or in combination with other risk factors (e.g., age, family history, race) or biomarkers in making decisions about which at risk patients to biopsy (KQ 1); (2) alone or in combination with other risk factors or biomarkers in making decisions about which at risk patients should consider repeat biopsy (KQ 2); and (3) alone or in combination with other biomarkers, risk factors or pathological markers in biopsy positive patients, in making decisions about aggressive treatment (e.g., radical surgery or radiation therapy) versus active surveillance (KQ 3). KQ 1 and KQ 2 reflect important clinical decision points. However, they are narrowly focused on a specific subset of patients (i.e., only initial or only repeat biopsy). Studies may not be limited to these two groups, or may not distinguish between the groups in their presentation of results in mixed populations. Therefore, it may be necessary to consider an alternative analytic approach to assess performance of the biomarkers in the two groups.

For prostate cancer prediction at biopsy or rebiopsy, the selected serum biomarker comparators were total prostate specific antigen (tPSA) elevations, free PSA (%fPSA), PSA density, PSA velocity or doubling time, and complexed PSA. Externally validated nomograms (EVNs), a type of risk assessment tool, were also reviewed. For classification as high or low risk,

initially selected comparators included Gleason score, tumor stage, other pathological tumor markers, prostate volume, and biomarkers used for risk classification (e.g., tPSA, fPSA, PSA density) and monitoring disease progression (e.g., tPSA, PSA velocity). The selected outcomes of interest for all KQs included both intermediate (diagnostic accuracy, impact on decisionmaking, harms of biopsy or treatment) and long-term (mortality, morbidity, quality of life) outcomes.

Methods

Methodological practices followed in this review were derived from AHRQ "Methods Guide for Effectiveness and Comparative Effectiveness Reviews"[70] (hereafter Methods Guide) and AHRQ "Methods Guide for Medical Test Reviews."[71]

Topic Development and Refinement

Key Questions (KQs) were reviewed and refined as needed by the Evidence-based Practice Center (EPC) with input from Key Informants and the Technical Expert Panel (TEP) to ensure that the questions were specific and explicit about what information was being reviewed. In addition, for Comparative Effectiveness Reviews, the KQs were posted for public comment and finalized by the EPC after review of the comments.

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the KQs for research that will inform health care decisions. The EPC solicits input from Key Informants when developing questions and an analytic framework for the systematic review or when identifying high priority research gaps and needed new research. The Key Informants selected to work on PCA3 included individuals with expertise in urology, pathology, laboratory medicine, internal medicine, family medicine, clinical trial design, as well as a patient advocate. The experts selected for the Technical Expert Panel to provide expertise and perspectives specific to the topic included individuals with expertise in urology, pathology, laboratory medicine, internal medicine, family medicine, clinical trial design and statistics. Technical Experts provided information to the EPC to identify literature search strategies and recommended approaches to specific issues as requested by the EPC.

Key Informants and Technical Experts are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism. Key Informants and Technical Experts were required to disclose any financial conflicts of interest greater than $10,000, and any other relevant business or professional conflicts of interest. Because of their role as end-users or their unique clinical or content expertise, individuals were invited to serve as Key Informants or Technical Experts, and those who presented without potential conflicts were retained. The AHRQ Task Order Officer (TOO) and the EPC worked to balance, manage, or mitigate any potential conflicts of interest identified.

Literature Search Strategy

The research librarian, in collaboration with the review team, developed and implemented search strategies designed to identify articles relevant to each KQ. Abstracts from selected recent professional meetings were also identified and followed up to identify subsequent publications and provide insight into types of data relevant to gaps in knowledge; abstract review was not used to assess publication bias. Details on strategies with full search strings are presented in Appendix A. The search was limited to English-language articles or articles in other languages for which the journal provided an English translation. The rational for this decision is that this EPC's experience demonstrated that non-English references did not yield information of sufficiently high quality to justify the resources needed for translation. In addition, studies have demonstrated that excluding non-English language studies has little impact on effect size

estimates or conclusions relative to the resources required.[72,73] Systematic reviews/meta-analyses were identified through the MEDLINE® searches and grey literature searches. Bibliographies of included articles were hand-searched to ensure complete identification of relevant studies. The timeframe for the search was limited to literature published after January 1, 1990 based on FDA approval of the tPSA test for early detection of prostate cancer in 1993.

- MEDLINE® (January 1, 1990 to August 9, 2011)
- Embase® (January 1, 1990 to August 15, 2011)
- Cochrane Central Register of Controlled Trials (no date restriction)

Search results were stored in a project-specific EndNote9® database that was subsequently uploaded into DistillerSR (Evidence Partners Inc., Ottawa, Ontario, Canada), a web-based systematic review software application. Two independent reviewers used the DistillerSR software to determine study eligibility. Using selection criteria for screening abstracts, the two reviewers marked each abstract as: 1) yes (eligible for review of the full article); 2) no (ineligible for review); or 3) uncertain (review the full article to resolve eligibility). Reviewer discrepancies were resolved by discussion and consensus opinion; a third reviewer was consulted as needed. When abstracts were unavailable or unclear, full-text articles were obtained for review.

Using study selection criteria and the DistillerSR software, a single reviewer read each full-text article and determined eligibility of the study for data abstraction. A second reviewer audited a subset of articles, and reviewed all articles marked as uncertain. Discrepancies were resolved by discussion and consensus opinion; a third reviewer was consulted as needed. Key reasons for excluding studies were captured by DistillerSR and Excel® spreadsheet. Each paper retrieved in full-text, but excluded from the review, is listed in Appendix B with reasons for exclusion.

An updated search of the published literature through May 15, 2012 was conducted upon submission of the draft report to determine if new information had been published since completion of the previous search (see Appendix A, Addendum). In addition, the Technical Expert Panel and individuals and organizations providing peer review were asked to inform the project team of any studies relevant to the KQs that were not included in the draft list of selected studies.

Study Selection

Studies were included if they fulfilled the following criteria:

- Study was a randomized controlled trial, a matched comparative study (e.g., prospective or retrospective cohort, diagnostic accuracy and case-control studies), or a systematic review of matched comparative studies. Matched studies were defined as performed in comparable clinical settings and provided test results and estimates of diagnostic performance for PCA3 and at least one other comparator (e.g., %fPSA) from the same patient population. A study of PCA3 alone, or a comparator alone, would not be included. Note that systematic reviews of unmatched studies were initially retained in DistillerSR (but not extracted) based on potential usefulness in two areas: 1) providing references that might identify additional studies of PCA3; and 2) as sources of more broadly based unmatched data on performance characteristics of PCA3 and comparators (i.e., to compare with results based on smaller numbers of subjects in the primary matched studies of %fPSA, to determine if the results are consistent or inconsistent).
- Study subjects were adult males with elevated total PSA tests and/or abnormal DRE who have not had a prostate biopsy or who have had one or more prostate biopsies (KQ 1 and 2), OR adult male patients with prostate cancer positive biopsies (KQ 3).

- Study intervention included testing for PCA3 and at least one designated pretreatment standard comparator test for prostate cancer, and a prostate biopsy (6 core minimum) or radical prostatectomy (KQ 3 only).
- Study comparators for KQ 1 and 2 were standard validated tests for prostate cancer that included tPSA, %fPSA, PSA velocity and doubling time, PSA density, complexed PSA and externally validated nomograms/risk assessment programs. For KQ 3, comparators included Gleason score, pathological staging, other pathological tumor characteristics and tumor volume.
- Study outcomes included intermediate outcomes (e.g., diagnostic accuracy for prostate cancer, impact on biopsy decisionmaking), long-term outcomes (e.g., mortality, morbidity, function, quality of life) and potential harms (e.g., adverse effects of biopsy, misdiagnosis) (KQ 1 and 2).

OR

- Study outcomes included the intermediate outcomes of diagnostic accuracy for tumor risk category (i.e., insignificant/low risk, aggressive/high risk) and impact on decisionmaking about active surveillance versus aggressive treatment, as well as long-term outcomes (e.g., mortality, morbidity, function, quality of life) and potential harms (e.g., adverse effects of treatment, misdiagnosis) (KQ 3).

Studies were excluded if they fulfilled at least one of the following criteria:

- Did not study prostate cancer.
- Did not address one or more of the KQs.
- Were published in a non-English language for which the journal did not provide a translation.
- Were published as a meeting abstract.
- Did not use a relevant study design.
- Did not report primary data.
- Did not report relevant outcomes.

Search Strategies for Grey Literature

A systematic search of grey literature sources was undertaken to identify unpublished studies, or studies published in journals that are not indexed in major bibliographic citation database, in accordance with guidance from Effective Health Care Scientific Resource Center. The detailed search strategies and results can be found in Appendix C. Briefly, the searches included: regulatory information (i.e., FDA); clinical registries; abstracts and papers from professional annual meetings and conferences; organizations publishing guidance or review documents (e.g., National Guideline Clearinghouse, Cochrane, National Institute for Clinical Excellence); grants and federally funded research; and manufacturer web sites.

Search strategies were similar to those used in bibliographic databases, except for the following:

- Regulatory information: The FDA website was searched for PMA and 510(k) decision summary documents related to urine PCA3 mRNA assays.
- For clinical registries, NIH RePORTER, HSRPROJ, and AHRQ GOLD, searches were limited to completed studies only.
- Abstracts and conferences articles published prior to 2009 were excluded.

Data Extraction and Management

The data elements from included studies were extracted using DistillerSR software into standard data formats and tables by one reviewer, and were subject to a full quality review for accuracy and completeness by a second reviewer. Data extraction question formats and tables were pilot-tested for completeness on a group of selected studies, and revised as necessary before full data extraction began. Project staff met regularly to discuss the results at each phase, review studies that were difficult to classify and/or abstract, and to address any questions raised by team members.

Data Elements

Data elements extracted from the selected studies were defined in consultation with the TEP. A detailed list can be found in Appendix B, and the corresponding database fields in the DistillerSR Data Extraction Forms in Appendix I.

Evidence Tables

DistillerSR reports were created that contained content for specific evidence tables and downloaded into Excel® spreadsheets for editing. Final tables were formatted in Microsoft Word®. Primary reporting of DistillerSR data elements for each evidence table was done by one person; a second person reviewed articles and evidence tables for accuracy. Disagreements were resolved by discussion and, if necessary, by consultation with a third reviewer. When small differences occurred in quantitative estimates of data from published figures, the values were obtained by averaging the two reviewers' estimates.

Individual Study Quality Assessment

Definition of Ratings for Individual Studies and Reviews

In adherence with the Methods Guide[70], grading the methodological quality of individual comparative studies was performed based on study design-specific criteria. In all cases, quality of individual studies and the overall body of evidence was assessed by two independent reviewers. Discordant decisions were resolved through discussion or third-party adjudication. Quality assessments were summarized for each study and recorded in tables. Criteria for assessing quality of nonrandomized comparative intervention studies and quality rating definitions[74,75] can be found in Appendix E. The quality of diagnostic accuracy studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool[76] that included the following 14 questions:

1. Was the spectrum of patients representative of the patients who will receive the test in practice?
2. Were the selection criteria clearly described?
3. Is the reference standard likely to classify the target condition correctly?
4. Is the period between the reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample receive verification by using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of the index test result?

7. Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)?
8. Was the execution of the index test described in sufficient detail to permit replication of the test?
9. Was the execution of the reference standard described in sufficient detail to permit replication of the reference standard?
10. Were the index test results interpreted without knowledge of the results of the reference standard?
11. Were the reference standard results interpreted without knowledge of the results of the index test?
12. Were the same clinical data available when the test results were interpreted as would be available when the test is used in practice?
13. Were uninterpretable/intermediate test results reported?
14. Were withdrawals from the study explained?

For KQ 1 and 2, the index test was PCA3 and the reference standard was biopsy. However, because selection of the screening positive populations was largely based on levels of tPSA, it was necessary to also consider QUADAS question 11 for tPSA to assess the potential for verification bias. This additional criterion was added at the end of the QUADAS questions (Table F-1, Appendix F), along with an entry to indicate whether verification bias was identified or suspected (response, Yes or No). Because measurement of specific clinical outcomes was needed to assess diagnostic accuracy for KQ 3, the additional criterion of clinical followup was added to the QUADAS questions (Table F-2, Appendix F).

The QUADAS ratings were summarized into general quality classes (from Paper 5, Table 5-4, AHRQ Test Review Guide[71]):

- **Good -** No major features that risk biased results.
- **Fair -** Susceptible to some bias, but flaws not sufficient to invalidate the results.
- **Poor -** Significant flaws that imply bias of various types which may invalidate the results.

Measuring Outcomes of Interest

There were several factors that supported the likelihood that most included studies would be focused on the predictive performance (e.g., clinical sensitivity and specificity, positive and negative predictive values for positive biopsy) of the PCA3 test. These factors included: the relatively short length of time that the PCA3 test, particularly the latest generation test, has been available; the comparative ease of conducting studies in which the end point is biopsy; and the length of time needed to collect long-term clinical outcomes related to the subsequent impact of interventions (e.g., active surveillance, treatment) related to the use of the test (as compared with no PCA3 testing or testing for other biomarkers). We expected that studies would provide a 2x2 table for PCA3 and other comparators, both for those subjects with positive biopsies and for those with negative biopsies (i.e., a matched analysis). In this way, one could evaluate not only the total performance of each test, but how the performance of the two tests varied in the population. For example, two tests could be shown to have equal sensitivity, but a matched analysis would indicate how often the two tests identified the same men with positive and negative biopsy results, and how often (and in what cases) they disagreed.

Two other intermediate outcomes for which data were sought were the impact of testing on physician and patient decisionmaking regarding biopsy and its potential harms (e.g., pain, bleeding, infection) and active surveillance versus treatment. Such data could be collected as

followup to biopsy via records review or by conducting surveys of physicians and patients. Use of surveys requires particular attention to uptake rates and the reliability, validity and disease-specificity of survey instruments.

Long-term outcomes or study endpoints (e.g., 7-15 years) of interest include mortality and survival, morbidity and clinical and biochemical failure.[3] All-cause mortality at different timeframes is reliable, but not a sensitive measure because it is dependent on age distribution and because most prostate cancer patients do not die of the disease. More sensitive measures are prostate cancer-specific 10-year survival or mortality if the cause of death is clear. Clinical failure may be measured as development of symptomatic disease, local disease progression or metastatic disease. Biochemical failure relates to increasing levels of total PSA (e.g., greater than 0.2 ng/mL) that may indicate disease recurrence. Morbidity also includes treatment-related adverse events (e.g., urinary incontinence, impotence) and other harms, as well as quality of life (QOL). Again, measuring QOL and the personal impact of symptoms related both to the cancer and to therapy requires the use of reliable and validated survey instruments. Minimally, assessment of QOL involves the use of a generic instrument to measure overall wellbeing, and a disease-specific instrument that focuses on specific symptoms and functions (e.g., incontinence, impotence).

Data Synthesis

After initial review of the extracted data from included studies, the analysis plan was finalized. No matched analyses were reported. However, only matched studies were included and pair-wise relative performance of PCA3 scores versus comparator results were summarized. Studies provided a wide variety of methods for comparing results, none of which were true matched analyses. For that reason, we chose to create the difference between the paired estimates and summarize these differences. Five separate analyses were designed:

1. A comparison of area under the ROC curve (or AUC);
2. Estimates of parameters defining the positive versus the negative biopsy populations;
3. Performance of PCA3 at a common cutoff score of 35;
4. Comparison of the ROC curves over a wide range of specificities/sensitivities; and
5. Results from logistic regression analysis.

As an example, consider a study reporting on a cohort of men age 50 or older who have a prostate biopsy and tPSA and PCA3 testing. For the first analysis, the AUC for tPSA was subtracted from the AUC for PCA3, resulting in the "difference of AUCs." This comparison is an unbiased estimate of effect size differences. The next retrieved study is analyzed in the same way, and the two differences are then compared for consistency across studies. This is repeated for all relevant studies, and then repeated for each of the five analyses. The entire process is then repeated for each comparator.

Due to the small number of relevant matched studies for most comparators, heterogeneity of results could only be explored for the PCA3/tPSA comparison. This included stratification by studies including men with all elevations of tPSA versus those focusing on the "grey zone" of borderline tPSA elevations. The analysis of tPSA was complicated by the presence of partial verification bias in all of the studies. We relied on published results and in-house modeling in an attempt to account for this bias, as original data were not available to use published correction methods.[77,78]

Modeling of PCA3 and tPSA performance could provide: 1) sensitivities of PCA3 and tPSA at set false positive rates and for a range of cutoffs, as well as a comparison of the number of

additional cases of prostate cancer detected by the better performing marker; and 2) specificities of PCA3 and tPSA at set sensitivities and for a range of cutoffs, as well as a comparison of the number of false positives avoided by the better performing marker. The model would need to be anchored by two important findings. First, that the ROC curves for tPSA (and for PCA3) were not influenced by the partial verification bias, and that PCA3 and tPSA are essentially independent markers. Sets of parameters (distribution descriptors such as means and standard deviations for PCA3 and comparators in both biopsy negative and positive men) derived from studies would need to fit the relevant ROC curve. A more detailed explanation of the methods used for performance modeling can be found in Appendix J. One aim of such modeling would be to more reliably explore the comparison of prostate cancer markers and assist in providing methods to more fully inform decision-making by men and their health care providers.

Based on the limited number of studies identified that addressed KQ 3, we anticipated focusing on a qualitative analysis (e.g., descriptive narrative, summary tables, identification of themes in content). Identification of more than one matched study in comparable populations, tested for PCA3 and one or more selected comparators, and reporting on the same intermediate or long-term clinical outcomes appeared to be unlikely.

Grading the Body of Evidence

The strength of evidence for primary outcomes was graded by using the standard process of the Evidence-based Practice Centers as outlined in the Methods Guide.[70] The method is based on a system developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group, and addresses four specific domains: risk of bias, consistency of effect sizes and direction of effect, directness of the link between evidence and health outcomes, and precision (or degree of certainty) of an effect estimate for a given outcome.[79] Additional domains (e.g., strength of association, dose-response relationship, plausible confounding, publication bias) can be assessed and reported if applicable based on the results of the evidence review. For this CER, grading was not limited to the KQs, but was applied to each outcome for each PCA3 comparator.

Based on the four required domains, each of these bodies of evidence was classified into one of four grade categories:

- **High -** High confidence that the evidence reflects the true effect. Further research is unlikely to change our confidence in the estimate of effect.
- **Moderate** – Moderate confidence that evidence reflects the true effect. Further research may change our confidence in the estimate of effect, or could change the estimate of effect.
- **Low -** Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate.
- **Insufficient –** Evidence either is unavailable or does not permit a conclusion.

The GRADE ratings were determined by independent reviewers, and disagreements were resolved by consensus as necessary.

Assessment of Applicability

Applicability of the results presented in this review was assessed in a systematic manner using the PICOTS framework (Population, Intervention, Comparison, Outcome, Timing,

Setting). Assessment included both the design and execution of the studies, and their relevance with regard to target populations, interventions and outcomes of interest.

Peer Review and Public Commentary

Peer reviewers and the public were invited to provide written comments on the draft report content based on their clinical and methodological expertise. Peer review comments on the preliminary draft were considered by the EPC in preparation of the final draft of the report. Peer reviewers did not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report did not necessarily represent the views of individual reviewers. The dispositions of the peer review comments will be documented and published three months after the publication of the evidence report. Potential reviewers were required to disclose any financial conflicts of interest greater than $10,000, and any other relevant business or professional conflicts of interest. Peer reviewers who disclosed potential business or professional conflicts of interest were able to submit comments on draft reports through the public comment mechanism.

Results

Literature Search

Of the 1,556 citations identified through the literature searches, 1,514 were excluded at various stages of review. One additional study was identified through grey literature searches. No additional studies were identified from one identified systematic review on PCA3[80]; this review was excluded from analyses as having no primary data. The 43 included articles reported the results of observational cohort studies with matched comparisons of PCA3 and other selected biomarkers. The PRISMA flow diagram (Figure 3) illustrates the review process for published studies, exclusions at each step and the selection results.

For Key Question (KQ) 1 and KQ 2, no randomized or comparative intervention trials were identified that included the use of PCA3 testing and reported long-term outcomes, or intermediate outcomes other than diagnostic accuracy. Of the 43 articles included, six studies addressing KQ 1 and KQ 2 were found to have duplicate data and were excluded from analyses[34,47,81-84]. Two other studies[85,86] were excluded because reported data were not in a format usable for these analyses. Of the remaining 34 studies, 24 addressed KQ 1 and/or KQ 2 (Table 1, Table 2a).

For KQ 3, no randomized or comparative intervention trials were identified that included the use of PCA3 testing and reported intermediate or long-term outcomes. Twelve observational studies were identified that addressed KQ 3 (Table 1, Table 2b), but one was excluded due to duplicate data.[87] Two studies reported on short-term health outcomes (i.e., biochemical recurrence and time to progression to treatment from active surveillance).[88,89] Table 1 provides general descriptive information on all studies. Table 2 describes study inclusion/exclusion criteria, and Table 3 describes the populations studied. Table 4 provides key information on PCA3 testing. In later sections, Table 17 details the characteristics of matched biopsy and prostatectomy studies addressing KQ 3; Table 18 provides information on comparators investigated along with PCA3 scores in studies addressing KQ 3.

Grey Literature Search

The process for evaluation of grey literature search results is summarized in Figure 4 and Appendix C. Two clinical trial registry citations were potentially relevant to the review:

- **Prostate Cancer Antigen 3 (PCA-3) Gene Project** (NCT01177436) – The status of this trial is unclear (last update August, 2010). Of interest was the use of three housekeeping genes for PCA3 testing in addition to *KLK3* (PSA): *ACTB* (beta-actin), *TUA* (Ka 1 tubulin), and *GAPDH* (glyceraldehyde-3-phosphate). Results may resolve remaining concerns about potential bias related to the use of *KLK3* as the housekeeping gene.
- **Clinical Evaluation of the Progensa® PCA3 Assay in Men With a Previous Negative Biopsy Result** (NCT01024959) – This trial, conducted by GenProbe and completed in April, 2011, provided data for the premarket (PMA) submission to FDA that was approved. However, the published article reporting the results of this clinical trial was not available for review.

Overall, the search of grey literature yielded one study, reported in the FDA Summary of Safety and Effectiveness Data for Gen-Probe's PROGENSA® PCA3 Assay (PMA P100033).

Figure 3. PRISMA diagram for selection of published studies

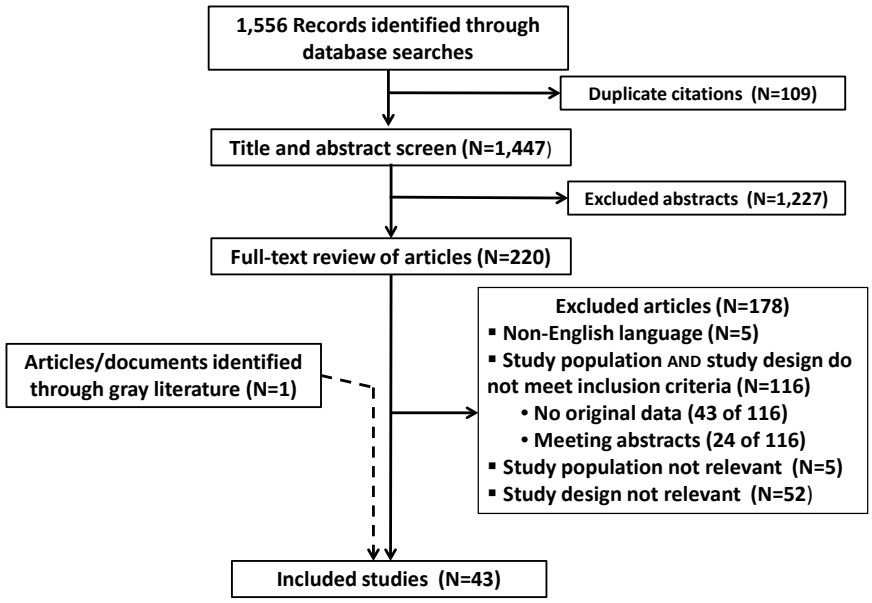
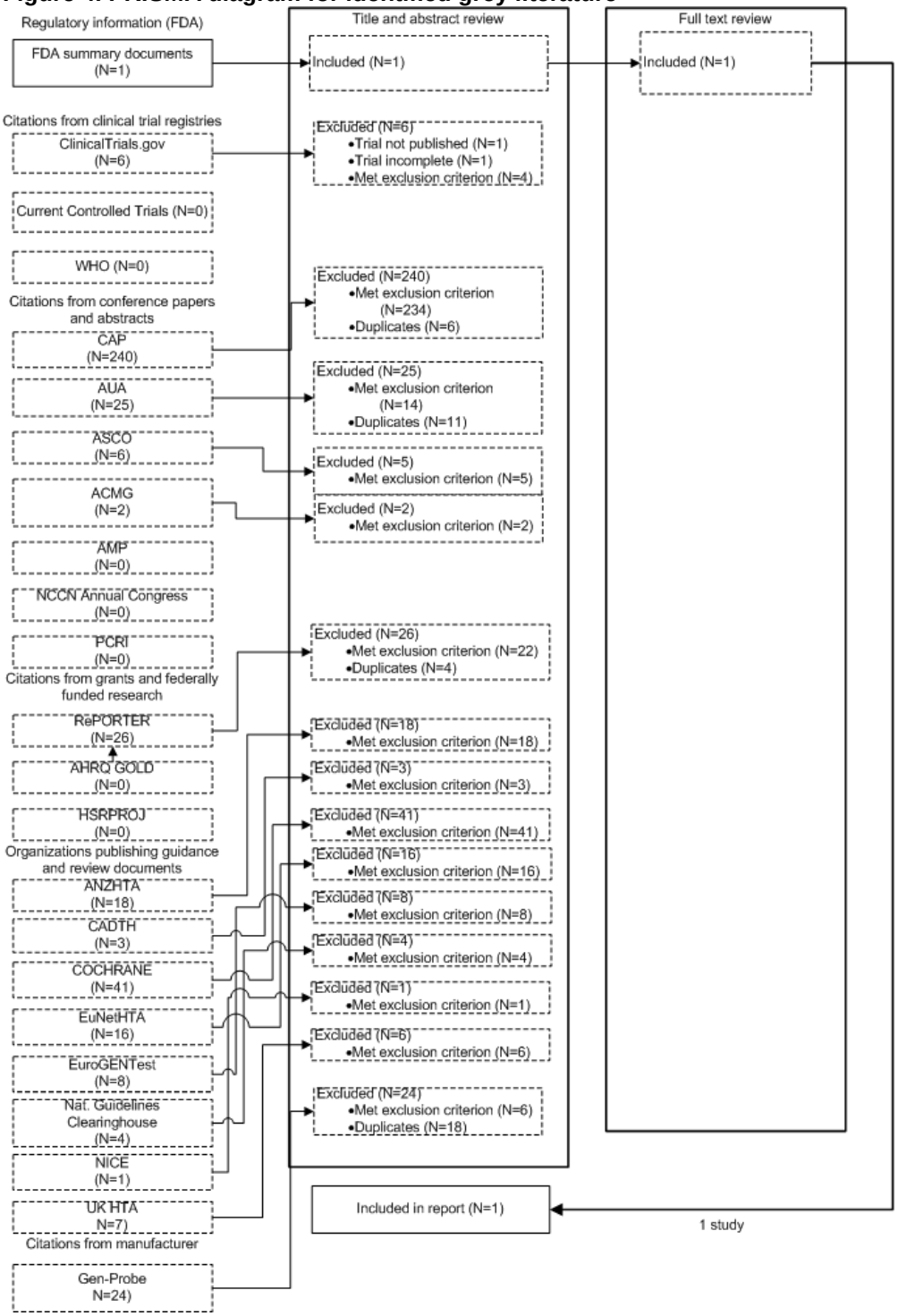


Figure 4. PRISMA diagram for identified grey literature



ACMG = American College of Medical Genetics; AHRQ = Agency for Healthcare Research and Quality; AMP = Association for Molecular Pathology; ASCO = American Society of Clinical Oncology; AUA = American Urological Association; CADTH = Canadian Agency for Drugs and Technologies in Health; CAP = College of American Pathologists; EuNetHTA = European Network for Health Technology Assessment; FDA = U.S. Food and Drug Administration; HSRPROJ = Health Services Research Projects in Progress; NCCN = National Cancer Consortium Network; NICE = U.K. National Institute for Health and Clinical Excellence; PCRI = Prostate Cancer Research Institute; RePORTER = National Institutes of Health Research Portfolio Online Reporting Tools Expenditures and Results tool; UK HTA = United Kingdom Health Technology Assessment Programme; WHO = World Health Organization

Table 1. General descriptions of the included studies in matched populations

Author[a], Year, Country	Enrollment Period (Month/Year)	Reference Standard	Biopsies Reported / Total Biopsies	Unreported Biopsies Explained[b]	% Positive Biopsies	Number P / RP	Key Question Addressed	Potential COI Funding[c] / Disclosures[d]
Adam[90], 2011, South Africa	07/09 - 02/10	Biopsy	105 / 107	Yes	43	--	1, 2	No / Yes
Ankerst[65], 2008 European cohort, NAm cohort	NR	Biopsy	443 / 443	NA	28	--	2	No / Yes
Aubin[91], 2010, International, U.S.	NR	Biopsy	1072 / 1140	Yes	18	--	2	Yes / Yes
Auprich[92], 2011 Austria, Germany, U.S.	7/08 – 7/09	Biopsy	127 / 127	NA	35	--	2	No / NR
Bollito[93], 2012, Italy	10/08 – 12/10	Biopsy	1237 / 1246	Yes	26	--	1, 2	No / No
Cao[94], 2011, China	06/09 – 04/10	Biopsy	131 / 143	Yes	60	--	1, 2	No / NR
de la Taille[95], 2011, Europe	02/08 - 08/09	Biopsy	516 / 528	Yes	40	--	1	Yes / Yes
Deras[96], 2008, Canada, U.S.	04/04 – 05/06	Biopsy	557 / 570	Yes	36	--	1, 2	NR / No
FDA Summary Document[97], 2012 U.S.	NR	Biopsy	466 / 495	Yes	22	--	2	Yes / Yes
Feero[98], 2012, Italy	5/10 – 12/10	Biopsy	151 / 151	NA	32	--	1	NR / NR
Goode[99], 2012, U.S.	NR	Biopsy	456 / 456	NA	19	--	1, 2	NR / NR
Hessels[100], 2010,The Netherlands	07/03 - 09/06	Biopsy/P	336 / 351	Yes	40	70	1, 2, 3	No / NR
Mearini[101], 2009, Italy	NR	Biopsy	96 / 96	NA	73	--	1, 2	No / No
Nyberg[102], 2010, Sweden	01/08 - 09/08	Biopsy	62 / 62	NA	29	--	1, 2	No / No
Ochiai,[103] 2011, Japan	05/07 - 05/08	Biopsy	105 / 105	Yes	36	--	1, 2	Yes / No
Ouyang[104], 2009, U.S.	NR	Biopsy	92/ 106	Yes	47	--	1, 2	No / NR
Pepe[105], 2012, Italy	10/09 – 09/11	Biopsy	74 / 74	NA	36	--	2	NR / NR
Perdona,[31] 2011, Italy	10/08 - 10/09	Biopsy	218 / 218	NA	33	--	1, 2	No / No
Ploussard[106], 2010, Europe	08/06 – 07/07	Biopsy	301 / 301	NA	24	--	2	NR / Yes
Rigau[107], 2010, Spain	NR	Biopsy	215 / 262	No	34	--	1, 2	No / NR
Roobol,[108]2010, Europe	09/07 - 02/09	Biopsy	721 / 721	NA	17	--	1, 2	NR / Yes
Schilling,[109] 2010, Germany	01/08 - 06/08	Biopsy	32 /32	Yes	56	--	1, 2	NR / No
Wang[110], 2009, U.S.	09/06 – 12/07	Biopsy	187 / 192	Yes	46	--	1, 2	No / Yes
Wu[111], 2012, U.S.	NR	Biopsy	103 / 188	Yes	36	--	2	No / NR
Auprich[112], 2011 Germany, U.S., Austria	11/06 - 10/09	P	--	--	--	305	3	No / No
Durand[113], 2012, France	02/09 – 06/10	P	--	--	--	160	3	NR / No

Table 1. General descriptions of the included studies in matched populations (continued)

Author[a], Year, Country	Enrollment Period (Month/Year)	Reference Standard	Biopsies Reported / Total Biopsies	Unreported Biopsies Explained[b]	% Positive Biopsies	Number P / RP	Key Question Addressed	Potential COI Funding[c] / Disclosures[d]
Kusuda[88], 2011, Japan	10/01 - 07/04	P	--	--	--	120	3	NR / NR
Liss[114], 2011, U.S.	05/07 - 04/08	P	--	--	--	98	3	Yes / NR
Nakanishi[115], 2008, U.S.	06/05 - 05/06	P	--	--	--	96	3	Yes / Yes
Ploussard[116], 2011, France	02/09 - 06/10	P	--	--	--	106	3	No / No
Tosoian[89], 2010, U.S.	00/95 – 00/09	Biopsy surveillance	294 / 301	No	13	--	3	Yes / NR
Van Poppel[117] 2011, Austria, Belgium, France, Netherlands	NR	P	--	--	--	175	3	Yes / Yes
Vlaeminck-Guillem[118], 2011, France	01/08 – 05/10	P	--	--	--	102	3	NR / Yes
Whitman[119], 2008, U.S.	09/06 - 11/07	P	--	--	--	72	3	Yes / Yes

COI = conflict of interest; FDA = U.S. Food and Drug Administration; NA = Not applicable; Nam = North American; NR = Not reported; P = prostatectomy; RP = radical prostatectomy

[a]In alphabetical order for studies with a reference standard of biopsy (N=24), then alphabetical order for studies with a reference standard of prostatectomy (N=9) or biopsy surveillance (N=1).

[b]'Unreported biopsies' are biopsies completed without results reported. This column indicates whether the studies provided an explanation for missing biopsy results.

[c]'Yes' indicates that funding for the study was provided entirely or in part by a developer of a PCA3 assay used in the study; 'No' indicates funding from another source.

[d]'Yes' indicates that one or more authors disclosed a paid consultancy or other relationship with the developer of a PCA3 assay used in the study.

Table 2a. Reported criteria for specific inclusion/exclusion of subjects for studies addressing KQ 1 and KQ 2

Author[a,b], Year	tPSA Cutoff (ng/ml)	Abnormal DRE (%)	Positive Biopsy	Initial Biopsy (%)	Positive Family History (%)	African-American	Other Risk Factors	Specific Exclusion Criteria
Adam[90], 2011	All values	Yes	No	82	Yes	Yes	NR	BPH, indwelling catheters
Ankerst[65], 2008	≥2.5	Yes (19)	No	0	5.9	NR	NR	NR
Aubin[91], 2010	2.5-10 (<60y) 3-10 (≥60y)	No	No	0	No	NR	NR	tPSA levels >10; meds that affect tPSA levels; HGPIN, ASAP
Auprich[92], 2011	≥2.5 to 6.5 - 50	Yes (11)	No	0	NR	NR	NR	tPSA levels > 50
Bollito[93], 2012	≥ 2.5	No	No	59	NR	NR	NR	Suspicious DRE; HGPIN; ASAP
Cao[94], 2011	≥4	Yes	No	NR	NR	NR	NR	Meds that affect tPSA levels
de la Taille[95], 2011	2.5-10	Yes (19)	No	100	NR	NR	NR	tPSA levels >10
Deras[96], 2008	≥ 2.5	Yes (15)	No	50	Yes	Yes	Yes	NR
FDA Summary[97], 2012	NR	Yes	No	0	NR	Yes	NR	Meds that affect tPSA levels; UTI / prostatitis; treatment
Feero[98], 2012	2.0-20	No	No	100	NR	NR	NR	tPSA levels > 20
Goode[99], 2012	≥ 4	Yes	No	63	NR	NR	NR	NR
Hessels[100], 2010	>3	Yes	No	Mixed; NR	Yes	NR	NR	NR
Mearini[101], 2009	>1	Yes	No	NR	NR	NR	NR	NR
Nyberg[102], 2010	≥ 2.5	Yes	No	55	NR	NR	NR	NR
Ochiai[103], 2011	2.5-50	Yes	No	81	NR	NR	NR	Meds that affect tPSA levels
Ouyang[104], 2009	≥ 4	NR	No	NR	NR	NR	NR	NR
Pepe[105], 2012	4-10	Yes	No	0	NR	NR	NR	tPSA levels > 10
Perdona[31], 2011	< 10	Yes (22)	No	61	No	No	NR	tPSA levels > 10; meds that affect tPSA levels; previous dx PCa
Ploussard[106]	2.5-10	No	No	0	NR	NR	NR	tPSA levels > 10; , meds that affect tPSA levels; initial biopsies
Rigau[107], 2010	≥ 4	Yes (26)	No	74	NR	NR	NR	Meds that affect tPSA levels
Roobol[108], 2010	≥ 3.0	No	No	71	No	NR	NR	NR
Schilling[109], 2010	>4	Yes	No	56	No	No	NR	NR
Wang[110], 2009	NR	Yes	No	73	Yes (19)	NR	Yes	History PCa
Wu[111], 2012	≥ 10	Yes (13)	No	0	NR	NR	NR	NR

Table 2b. Reported criteria for specific inclusion/exclusion of subjects for studies addressing KQ 3

Author[a,b], Year	Positive Biopsy	Description	Specific Exclusion Criteria
Auprich[112], 2011	Yes	Clinically localized PrC	Meds that affect PSA levels
Durand[113], 2012	Yes	Localized PrC	NR
Hessels[100], 2010	Yes	PrC	NR
Kusuda[88], 2011	Yes	Clinically localized PrC	NR
Liss[114], 2011	Yes	PrC	NR
Nakanishi[115], 2008	Yes	PrC	Not age 40-70, PSA levels ≥ 50; or taking meds that affect PSA levels
Ploussard[116], 2011	Yes	"Low risk" localized PrC	Not "low risk" (PSA ≤ 10; stage T1c-T2a; Gleason score 6)
Tosoian[89], 2010	Yes	"Low risk" PrC patients (Epstein criteria) in a surveillance program	Positive biopsy but not "low risk" based on Epstein criteria
van Poppel[117], 2011	Yes	Biopsy positive PrC	NR
Vlaeminck-Guillem[118], 2011	Yes	Biopsy positive	Neoadjuvant treatment
Whitman[119], 2008	Yes	Biopsy positive	Meds that affect PSA levels

ASAP = atypical small acinar proliferation; BPH = benign prostatic hyperplasia; DRE = digital rectal exam; HGPIN = high-grade prostatic intraepithelial neoplasia; NA = Not applicable; No = Not a criterion for inclusion; NR = not reported; PrC = prostate cancer; PSA = prostate specific antigen; tPSA = total prostate specific antigen; UTI = urinary tract infection

[a]In alphabetical order for KQ 1/KQ 2 studies (N=24) in Table 2a, then in alphabetical order for KQ 3 studies (N=11) in Table 2b.

[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients. In Table 2a, rows 3, 7, 10, and 17-19 are shaded. There is no shading in Table 2b.

Table 3. Characteristics of populations studied

Author[a], Year	Reference Standard	Age Distribution in Years Mean (sd) / Median (Range)	Race Distribution, %	% Positive DRE	% Positive Family History	Coexisting Pathology Name, %
Adam[90], 2011	Biopsy	NR / 67 (35-89)	White 25.7; AA 68.6; Other 5.7	48.6	4.8	NR
Ankerst[65], 2008	Biopsy	NR / 66 (11-83)	White 97.5; AA 2.0; Other 0.5	18.7	NR	NR
Aubin[91], 2010	Biopsy	66.1 (6.0) / 66.1 (52.7-80)	NR	NR	13.8	NR
Auprich[92],2011	Biopsy	NR / 63 (50-70)	NR	11	NR	NR
Auprich[112],2011	P	NR / 63 (44-79)	NR	23.9	NR	NR
Bollito[93], 2012	Biopsy	NR / 67 (42-89)	NR	0	NR	NR
Cao[94], 2011	Biopsy	NR	Asian (Chinese) 100	NR	NR	NR
de la Taille[95], 2011	Biopsy	63 (7.6) / 63	NR	NR	NR	NR
Deras[96], 2008	Biopsy	64 / 64 (32-89)	White 82.5; AA 5.3; Hispanic 2.3; Asian 0.4; Other 9.6	NR	NR	NR
FDA Summary[97], 2012	Biopsy	NR / 67 (44-92)	White 87.5; AA 9.1; Other 2.6	NR	NR	NR
Feero[98], 2012	Biopsy	NR / 64 (48-87)	NR	0	NR	HGPIN, 24; BPH, 44
Goode[99], 2012	Biopsy	NR / 66 (41-90)	NR	NR	NR	HGPIN; ASAP
Hessels[100], 2010	Biopsy	63 / 64 (38-83)	NR	NR	NR	NR
Kusuda[88], 2011	P	67.2 (6.5) / NR	NR	NR	NR	NR
Liss[114], 2011	P	62.7 (7.2) / NR	NR	16	NR	NR
Mearini[101], 2009	Biopsy	NR	NR	NR	NR	BPH, 27
Nakanishi[115], 2008	P	60 / 60 (45-70)	White 78.1; AA 15.6; Hispanic 6.3	NR	NR	NR
Nyberg[102], 2010	Biopsy	NR / 63 (IQR 57-70)	NR	NR	NR	HGPIN, 5
Ochiai[103], 2011	Biopsy	NR / 66 (44-87)	Asian (Japanese) 100	NR	NR	NR
Ouyang[104], 2009	Biopsy	NR	White 98; AA 2	NR	NR	NR
Pepe[105], 2012	Biopsy	NR / 64 (48-74)	NR	NR	NR	NR
Perdona[31], 2011	Biopsy	NR / 66 (60-72)	NR	NR	4.6	NR
Ploussard[106], 2010	Biopsy	64.6 / 65.4 (43.3-83.4)	NR	NR	NR	ASAP, 4
Ploussard[116], 2011	P	62 / 62 (43-75)	NR	NR	NR	NR
Rigau[107], 2010	Biopsy	Mean 65.7 / Range 44-85	NR	26.5	NR	NR
Roobol[108], 2010	Biopsy	70.0 / 70.2 (63.6-77.5)	NR	13.1	NR	NR

Table 3. Characteristics of populations studied (continued)

Author[a], Year	Reference Standard	Age Distribution in Years Mean (sd) / Median (Range)	Race Distribution, %	% Positive DRE	% Positive Family History	Coexisting Pathology Name, %
Schilling[109], 2010	Biopsy	NR / 67 (42-88)	NR	NR	NR	NR
Tosoian[89], 2010	Biopsy surveillance	68.2 (6.2) / 68.2 (50.3-84.2)	White 91.8; AA 5.4; Other 2.8	NR	NR	NR
Vlaeminck-Guillem[118], 2011	P	62 (6) / 63 (47-72)	NR	NR	NR	NR
Wang[110],2009	Biopsy	62 (8.3) / Median NR (44-86)	White 91.5; AA 5.3; Other 3.2	16	18.7	HGPIN, 5.9 ASAP, 2.1
Whitman[119], 2008	P	NR / 58 (42-73)	White 75; AA 25	NR	NR	NR
Wu[111], 2012	Biopsy	63.5 (7.4) / NR	NR	13	NR	HGPIN; ASAP

AA = African American; ASAP = atypical small acinar proliferation; BPH = benign prostatic hyperplasia; DRE = digital rectal exam; HGPIN = high-grade prostatic intraepithelial neoplasia; sd = standard deviation; tPSA = total prostate specific antigenNR = not reported; P = prostatectomy

[a]Shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients. Rows 3, 8, 11, and 21-23 are shaded.

Table 4. PCA3/DD3 assay characteristics

Author[a], Year	Attentive Massage Used	Specimen	Method Used	Assay Specified	Reporting Units	House-keeping Gene	Handling Temperatures Holding (C); Storage (C)	Informative Results %
Adam[90], 2011	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	98
Aubin[91], 2010	Yes	UU	TMA	"Gen-Probe"	PCA3 Score	PSA	NR; -70	94
Auprich[92], 2011	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	NR
Auprich[112], 2011	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Bollito[93], 2012	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	96
de la Taille[95], 2011	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	99
Deras[96], 2008	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR; -70	100
FDA Summary Document[97], 2012	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	93
Feero[98], 2012	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	NR
Goode[99], 2012	Yes	UU	TMA	Progensa reagents by Bostwick	PCA3 Score	PSA	NR	100
Hessels[100], 2010	Yes	US	TMA	Aptima	PCA3 Score	PSA	NR	96
Liss[114], 2011	Yes	UU	TMA	NR	PCA3 Score	PSA	2-8; NR	100
Nakanishi[115], 2008	Yes	UU	TMA	"Gen-Probe"	PCA3 Score	PSA	2-8; -70	100
Nyberg[102], 2010	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Ochiai[103], 2011	Yes	UU	TMA	Aptima	PCA3 Score	PSA	NR; -70	100
Perdona[31], 2011	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Ploussard[116], 2011	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	NR
Roobol[108], 2010	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Schilling[109], 2010	Yes	UU	TMA	"Gen-Probe"	PCA3 Score	PSA	NR	99
Tosoian[89], 2010	Yes	UU	TMA	Aptima	PCA3 Score	PSA	NR; -80	98
Vlaeminck-Guillem[118], 2011	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Wang[110], 2009	Yes	UU	TMA	"Gen-Probe"	PCA3 Score	PSA	30; NR	97
Whitman[119], 2008	Yes	UU	TMA	"Gen-Probe"	PCA3 Score	PSA	NR	100

Table 4. PCA3/DD3 assay characteristics (continued)

Author[a], Year	Attentive Massage Used	Specimen	Method Used	Assay Specified	Reporting Units	House-keeping Gene	Handling Temperatures Holding (C); Storage (C)	Informative Results %
Fradet[85], 2004	Yes	UU	NASBA	uPM3	Probability	PSA	NR; 2-8	86
Tinzl[86], 2004	Yes	UU	NASBA	uPM3	PCA3 Score	PSA	4; -20	79
Wu[111], 2012	Yes	UU	TMA	PCA3*Plus* by Bostwick	PCA3 Score	PSA	NR	NR
Cao[94], 2011	Yes	US	QRT-PCR	NR	PCA3 Score	PSA	4; -20	92
Kusuda[88], 2011	NR	LN	QRT-PCR	NR	Quantification	GADPH	NR	NR
Mearini[101], 2009	Yes	US	QRT-PCR	NR	Quantification	Beta actin	NR -70	NR
Ouyang[104], 2009	Yes	US	QRT-PCR	NR	PCA3 Score	GADPH	NR; -80	87
Rigau[107], 2010	Yes	US	QRT-PCR	NR	PCA3 Score	PSA	NR; -80	82
Ankerst[65], 2008	NR	NR	NR	NR[b]	PCA3 Score	PSA	NR	NR
Ploussard[106], 2010	NR	UU	NR	NR[c]	PCA3 Score	PSA	NR	NR

LN = lymph node tissue; NASBA = nucleic acid sequence-based amplification; NR = not reported; QRTPCR = quantitative real time polymerase chain reaction; TMA = transcription-mediated amplification; U = urine unsedimented; US = urine sedimented.

[a]Studies are in alphabetical order by methodology.

[b]Ankerst analyzed data from a study concurrently reported by Haese et al., 2008[82]. Haese indicated that PCA3 testing was done using Progensa.

[c]This paper did not specify the PCA3 assay used, but two authors (van Poppel, de la Taille) disclosed an advisor and investigator relationship with GenProbe, making it likely that Progensa was the assay used.

Potential Biases in Included Studies

The populations in the included studies were largely drawn from academic medical centers where patients with elevated tPSA results and/or other risk factors (e.g., positive DRE, family history, African American race) were seeking referral or specialty care. Observational studies of such opportunistic cohorts are subject to specific biases.

Verification Bias

Men will be offered prostate biopsy based on the extent of tPSA elevations, suspicious findings on a digital rectal exam (DRE), a combination of the two or, less commonly, other risk factors such as family history or race. In order to obtain an unbiased estimate of diagnostic accuracy for tPSA at specific cutoffs, it is necessary that the identification of prostate cancer not be related to tPSA levels. This is a potential problem as studies have shown that higher tPSA levels are indicative of a higher likelihood for the presence of prostate cancer. Men are more likely to undergo prostate biopsy, if the tPSA is high (e.g., 10-20 ng/mL), rather than close to lower cutoffs used to define a positive tPSA screening test (e.g., 3-4 ng/mL). If a study reports results in which biopsy is tPSA-related, the sensitivity and specificity at select tPSA cutoffs will not be accurate. If those not accepting biopsy are considered missing, this is considered "partial verification" bias. All studies included in the evidence review are opportunistic cohorts of men agreeing to biopsy, and will be subject to this bias. However, no study addressed this potential bias. We addressed this bias through modeling the effect of verification bias on the tPSA measurements. A detailed discussion of this bias and the modeling performed can be found in Appendix J.

Spectrum Bias

Spectrum effects should also be considered when evaluating diagnostic tests generated from convenience samples collected at referral centers. For example, such studies are likely to represent men at higher risk of prostate cancer than in the total cohort of screened men, and might be at higher risk of more aggressive cancers as well (e.g., those with rapid rise in PSA). The positive biopsy rate in such referral populations will depend on multiple factors, including the tPSA cutoff, the number of men with elevated tPSA who opt out of biopsy (e.g., men with lower tPSA levels and lower risk), and/or the proportions of men with other important risk factors. In 17 included studies, biopsy positive rates ranged from 16.9 to 72.9 percent, with a median of 36 percent. Although this may not influence the clinical sensitivity and specificity estimates it certainly will influence the positive and negative predictive values (as the disease prevalence varies).

A second spectrum effect of more concern relates to the range of severity of disease between those identified with an elevated PCA3 score compared with those with a positive comparator test. For example, suppose two tests have the same sensitivity and specificity estimated in a cohort of biopsied men. In order to show the true clinical validity, it would be necessary to examine the men with positive biopsies having discordant test results (i.e., positive by one test but negative by the other). If the men positive by one test have similar tumor characteristics and/or severity of disease to those positive by the other test, then the sensitivity/specificity estimates could be both statistically and clinically equivalent. However, if one test identifies a

difference in tumor characteristics and/or severity of disease that the other test does not, the estimates could be statistically equivalent but clinically different.

Sampling Bias

Analysis of tPSA (and related comparators) was also subject to a sampling bias. A subset of studies restricted enrollment to tPSA results in the "grey zone" (e.g., 2.5 ng/mL to less than 10 ng/mL). The impact of this restriction would be to reduce the prevalence of disease in the study group, as there is a positive correlation between tPSA and prevalence of prostate cancer. In addition this would also reduce the tPSA test performance (sensitivity/specificity) as those men with higher tPSA levels are not enrolled in the study. It is at the higher tPSA levels that the test is most predictive. This would reduce the apparent performance of tPSA measurements, and increase the difference between the PCA3 and tPSA performance estimates. Overall, this bias reduces the external validity (generalizability) of studies in a general population. This bias cannot be avoided by statistical analysis. These studies could be removed from consideration. Instead, we have chosen to address this bias by stratifying analyses by selection criteria. That is, studies of the "grey zone" were summarized separately from the studies that include all levels of elevated tPSA.

Analyses Relating to KQs 1 and 2

KQ 1: Testing PCA3 and Comparators To Identify Prostate Cancer in Men Having an Initial Biopsy

Among the 17 studies addressing KQ 1 (Table 1), only two reported results in populations where all men were having initial biopsies (Table 5).[95,98] Both studies reported data on tPSA and %fPSA; one[95] also reported on PSA density. The actual data will be presented in later analyses (KQ 1/KQ 2 combined), but there were too few data for reliable interpretations. The five matched analyses performed (see Table 5 footnote a) are outlined in Methods, and discussed in detail as part of the analyses.

Table 5. Summary of matched analyses (of five)[a] that can be performed for each PCA3 comparator, with analysis restricted to studies only on initial biopsies

Study Author[b], Year	N	Initial Biopsy	tPSA	%fPSA	PSA Velocity	PSA Density	cPSA	Validated Nomogram
de la Taille[95], 2011	516	100%	A,C,D,E	C,D,E	-	A,C,D	-	-
Ferro[98], 2012	151	100%	B	A,D	-	-	-	-
All	**667**							

%fPSA = percent free prostate specific antigen; cPSA = complexed prostate specific antigen; N = number; tPSA = total prostate specific antigen

The dash ('-') indicates no data provided for that comparator.

[a]The letters 'A' through 'E' represent the analyses: A = AUC, B = Mean/median values, C = PCA3 > 35, D = Sensitivity/Specificity and E = Regression.

[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients. Both rows are shaded.

Strength of Evidence

When data were restricted to the two studies reporting only on populations of men having an initial prostate biopsy, only one comparison could be made, the D analysis for %fPSA. Both

studies were poor quality. It is not possible to evaluate consistency (between-study results). In addition, estimates of effect size will be imprecise. This results in assigning grades of "insufficient" for the reported comparisons of PCA3 with tPSA, %fPSA, and PSA density, and other comparators (PSA velocity, complexed PSA and externally validated nomograms) with no matched studies.

KQ 2: Testing PCA3 and Comparators in Men Having Repeat Biopsy

Among the 21 studies addressing KQ 2 (Table 1, Table 2b), seven[65,91,92,97,105,106,111] reported results in populations where all of the men were having a repeat biopsy (Table 6). Studies are ranked by number of patients enrolled. Five studies reported on tPSA,[65,91,92,106,111] four on %fPSA[91,92,105,106] and two on externally validated nomograms.[65,97] All studies were poor quality. The actual data were included in later analyses, but there were too few data for any one analysis to provide reliable interpretations.

Table 6. Summary of matched analyses (of five)[a] that can be performed for each PCA3 comparator, with analysis restricted to studies reporting results only in repeat biopsy

Study Author[b], Year	N	Initial Biopsy	tPSA	%fPSA	PSA Velocity	PSA Density	cPSA	Validated Nomogram
Pepe[105], 2012	74	0%	-	B[a]	-	-	-	-
Wu[111], 2012	103	0%	A,D	-	-	C	-	-
Auprich[92], 2011	127	0%	A,B	B,D	B,D	-	-	-
Ploussard[106], 2010	301	0%	A,D	A,D	-	-	-	-
Ankerst[65], 2008	443	0%	A,B,D	-	-	-	-	A
FDA Summary[97], 2012	466	0%	-	-	-	-	-	A,E
Aubin[91], 2010	1,072	0%	C,D	A,D	-	-	-	-
All	**2,586**							

%fPSA = % free prostate specific antigen; cPSA = complexed prostate specific antigen; FDA = U.S. Food and Drug Administration; N = number; tPSA = total prostate specific antigen

A dash ('-') indicates no data provided for that comparator.

[a]The letters 'A' through 'E' represent the analyses: A = AUC, B = mean/sd, C = PCA3 > 35, D = Sensitivity/Specificity and E = Regression.

[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients. Rows 1, 4, and 7 are shaded.

Strength of Evidence

When data were restricted to these seven studies, the number of comparisons possible for each matched analysis remained small due to three "grey zone" studies.[91,105,106] For example, the "D" analysis for tPSA has three sets of data, but two are for all levels of tPSA and one is restricted to the "grey zone." All studies were poor quality. Due to these differences in inclusion criteria, it is difficult to evaluate consistency. Estimates of effect size will, necessarily, also be imprecise. Strength of evidence was deemed insufficient for all comparisons of PCA3 with tPSA, %fPSA, PSA velocity, PSA density, complexed PSA and externally validated nomograms in this population of men.

Potential To Combine KQs 1 and 2: Testing PCA3 and Comparators in Men Having Initial or Repeat Biopsy

The sections above addressed the nine studies that exclusively studied men having an initial (KQ 1) or repeat biopsy (KQ 2). However, 15 additional studies included matched results of PCA3 and the comparators (Table 7). Eleven reported the proportion of men having initial and repeat biopsies,[31,90,93,96,99,102,103,107-110] but four did not report biopsy history.[94,100,101,104] The results from these studies were most often not stratified by biopsy history.

At this point, one could have ignored the data in these 15 additional studies, as they did not directly apply to either KQ 1 or KQ 2. Instead, based on the inadequate strength of evidence found for the prior individual analyses that focused on only those men with initial or with repeat biopsies, we chose to examine whether data from the studies that could be stratified by biopsy history might be suitable for a combined analysis (Table 7). Prior to performing this combined analysis, however, it was necessary to determine whether biopsy status was an important covariate that could bias the findings. An examination of Table 7 found that the most common comparator was tPSA, and the most common analysis, by far, was the area under the curve (AUC), indicated by an "A."

Fifteen of the 19 studies that reported AUC results for both PCA3 and tPSA also provided the proportion of study subjects with no previous prostate biopsies. A regression analysis of AUC difference (PCA3 – tPSA) versus the proportion of men with an initial biopsy would provide evidence regarding suitability of the combined analysis. The raw data for this figure can be found in Table 10.

Figure 5A shows the analysis. Based on linear regression, the slope (-0.00227) was not significant (p=0.97), indicating that there was no significant relationship between the biopsy status and AUC difference for PCA3 versus tPSA elevations. In addition, a subset of three of the 15 studies reported AUCs stratified by initial and repeat biopsy status (Table 10). Figure 5B shows the analysis with the replacement of the three "composite" AUCs with initial and repeat biopsy subgroup AUCs. The slope (-0.01307) was also not significant (p=0.81).

Examining Table 7 also indicated that 14 studies (16 datasets) reported the ROC curves for both PCA3 and tPSA. A regression analysis of (PCA3 – tPSA) sensitivities at a constant specificity of 50 percent versus the proportion of men with an initial biopsy would also provide evidence regarding the suitability of the combined analysis. The raw data for this figure can be found in Table 13. Figure 6 shows the described analysis. Based on linear regression, the slope (0.02956) is not significant (p=0.79). Again, there appears to be little or no association between the biopsy history and the relative performance of PCA3 and tPSA.

Together, these two analyses shown in Figures 5 and 6 provided evidence that combining results from studies of initial biopsies, repeat biopsies, and mixtures of initial and repeat biopsies did not appear to impact the comparison of PCA3 with tPSA elevations.

Table 7. Summary of matched analyses (of five)[a] that can be performed for each PCA3 comparator, with analysis restricted to studies on initial and repeat biopsies

Author Study[b], Year	N	Initial Biopsy	tPSA	%fPSA	PSA Velocity	PSA Density	cPSA	Validated Nomogram
Pepe[105], 2012	74	0%	-	B[a]	-	-	-	-
Wu[111], 2012	103	0%	A,D	-	-	A,C,D	-	-
Auprich[92], 2011	127	0%	A,B	A,B	A,B	-	-	-
Ploussard[106], 2010	301	0%	A,D	A,D	-	-	-	-
Ankerst[65], 2008	443	0%	A,B,D	-	-	-	-	A
FDA Summary[97], 2012	446	0%	-	-	-	-	-	A,E
Aubin[91], 2010	1,072	0%	C,D	A,D,E	-	-	-	-
Deras[96],2008	557	51%	A,C,D	-	-	-	-	-
Nyberg[102], 2010	62	55%	A,B,C,D	-	-	-	-	-
Bollito[93], 2012	1246	59%	A,B,D	A,B,D	-	-	-	-
Perdona[31], 2011	218	61%	A,B,D,E	B,E	-	-	-	A,B,D
Goode[99], 2012	456	63%	A,C,D	-	-	D	-	-
Roobol[108], 2010	721	71%	A	-	-	-	-	-
Wang[110], 2009	181	73%	C	-	-	-	-	-
Rigau[107]2010	215	74%	A,D	-	-	-	-	-
Ochiai[103], 2011	105	81%	A,B,C,D	-	-	A,B,D	-	D
Adam[90], 2011	105	82%	A,C,D	-	-	-	-	-
Schilling[109], 2010	32	86%	A,C,D	-	-	-	-	-
Ferro[98], 2012	151	100%	B	A,D	-	-	-	-
de la Taille[95], 2011	516	100%	A,C,D,E	C,D,E	-	A,C,D	-	-
Ouyang, 2009[104]	92	-	A	-	-	-	-	-
Mearini[101], 2009	96	-	A	-	-	-	-	-
Cao[94], 2011	131	-	A	-	-	-	-	-
Hessels[100], 2010	336	-	A	-	-	-	-	-

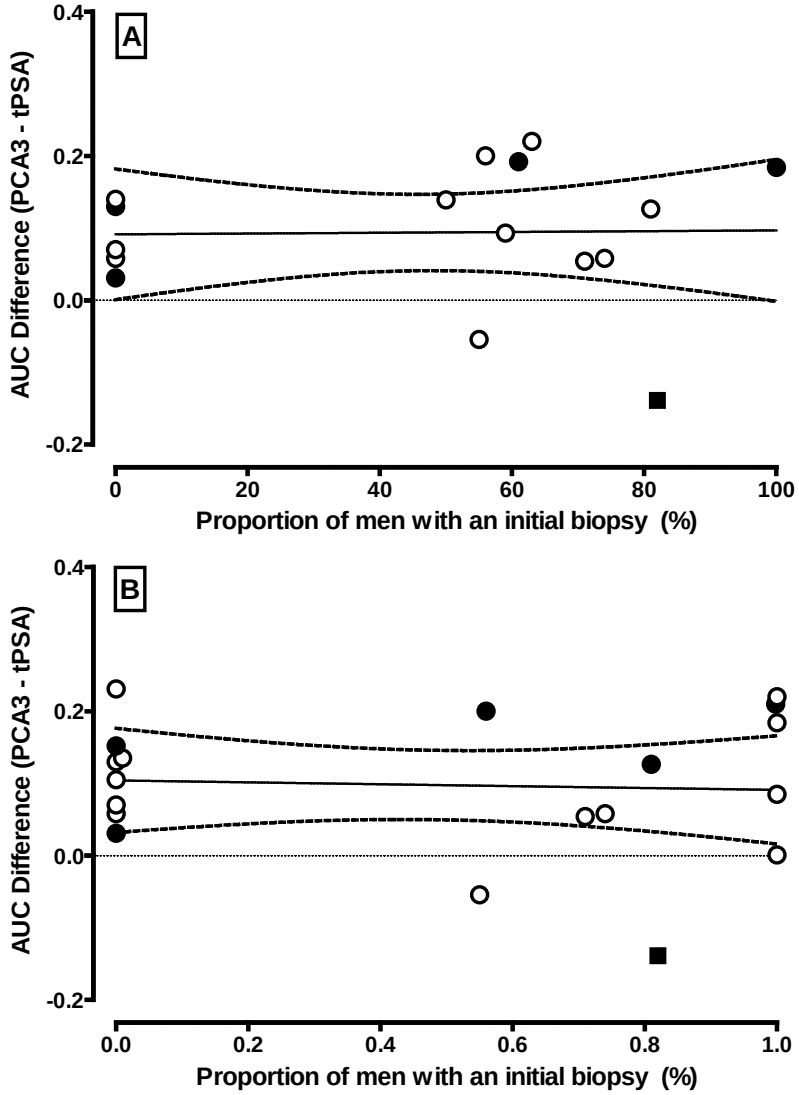
%fPSA = % free prostate specific antigen; cPSA = complexed prostate specific antigen; N = number; tPSA = total prostate specific antigen

A dash ('-') indicates no data provided for that comparator.

[a]The letters 'A' through 'E' represent the analyses: A = AUC, B = mean/sd, C = PCA3 > 35, D = Sensitivity/Specificity and E = Regression.

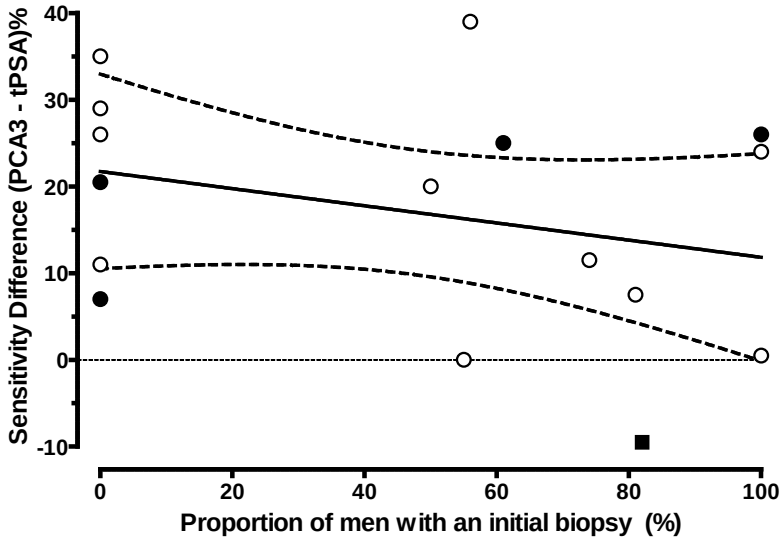
[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (2.5-10 ng/mL) when enrolling patients. Rows 1, 4, 7, 11, 19, and 20 are shaded.

Figure 5. Scatterplots showing the relationship between the proportion of men with an initial biopsy versus the differences in the area under the curves (AUC) for PCA3 scores versus tPSA elevations



AUC = area under the curve; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
Note: The horizontal axis shows the proportion of men with an initial biopsy (range 0 to 100%). The vertical axis shows the AUC Difference ($AUC_{PCA3} - AUC_{tPSA}$); a value of 0 (dashed horizontal line) indicates the two AUCs are equivalent. Each circle indicates the results from one included study. Results from studies reporting only on the "grey zone" (e.g., tPSA results 2.5-10 ng/mL) are filled. Figure 5A shows studies providing the proportion of men with initial biopsies as presented in Tables 5 and 7. There is no significant relationship, as indicated by the solid line and 95% CI. Figure 5B draws data from the same studies, but includes data for the initial and repeat biopsy groups from three studies[31,93,99] summarized as composite data in Figure 5A. There is still no significant relationship. The most outlying study had a negative AUC difference and reported data from a 69% black South African study (82% initial biopsy).[90]

Figure 6. A scatterplot showing the relationship between the proportion of men with an initial biopsy versus the difference in PCA3 and tPSA sensitivities at a specificity of 50%



PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
Note: The horizontal axis shows the proportion of men with an initial biopsy (range 0 to 100%). The vertical axis shows the difference in sensitivity (PCA3 sensitivity – tPSA sensitivity) of detecting prostate cancer, when the specificity is held constant at 50%; a value of 0 (dashed horizontal line) indicates the sensitivities are equivalent. Each circle indicates the results from one included study. Results from four studies focusing on the "grey zone" [31,91,95,106] are filled circles. The original data for this figure were shown in Table 13. There was no significant relationship (slope = -0.09, p=0.26) for the 16 studies, as indicated by the solid line and 95% confidence intervals. Removing the one study[90] (filled square) from South Africa in which 69% of men were black resulted in a slope closer to zero (slope = -0.06; p=0.44). Removing this study and the four "grey zone" studies results in a similar slope (slope = -0.15, p=0.16).

One other matched analysis provided additional support for this finding. Two studies[93,99] provided matched PCA3/tPSA ROC curves separately for men who had an initial biopsy, and those receiving a repeat biopsy (also Table 13). Therefore, these were matched within-study comparisons of the relative effectiveness of the two markers in these biopsy-specific subgroups. The first study[93] found that tPSA performed much better in initial compared with repeat biopsies, while PCA3 performed much worse in initial compared with repeat biopsied men. This is consistent with some who have argued that tPSA would be expected to perform poorly in the repeat biopsy setting, as those tumors associated with high tPSA were identified in the initial round of testing and would not be present in a population having repeat biopsies. However, the second study[99] found much different patterns. The tPSA performed almost equally as well in initial and repeat biopsy settings, and PCA3 performed much better in initial compared with repeat biopsies. These two studies reported almost opposite findings.

Such analyses cannot be performed for any of the other comparators. However, given the lack of data for those comparisons, we chose to comprehensively list all potentially relevant results, regardless of the biopsy status of the enrolled men. The following sections provide the results of the combined analysis of KQ 1 and KQ 2.

Description of Included Studies for KQ 1/KQ 2 "Combined"

As noted in Methods, the inclusion criteria restricted study inclusion to matched studies. These were defined as studies that provided estimates of diagnostic test performance for PCA3 and at least one other comparator (e.g., tPSA elevations or %fPSA) using the same patient population. Thus, a study of PCA3 alone, or a comparator alone, would not be included. In examining the included studies, it was clear that, although the same population was used, the reports rarely applied a true matched analysis. However, the results were still considered as being "matched," due to the application of the test(s) to the same underlying population. We preserved this population matching by computing differences between PCA3 test results and comparator test results within each study. These matched differences could then be compared across studies.

For example, one study of biopsied patients might report an AUC for PCA3 and then separately report an AUC for tPSA in the same population. The difference in the two would then be computed and compared with the difference in AUCs from other similarly matched studies. Although this restriction limited the number of included publications, it was aimed at improving the consistency of results. For example, an analysis of unmatched studies might have provided sufficient information to stratify PCA3 performance by number of previous biopsies. Similar data could be obtained from the literature for tPSA. Comparing the results between these unmatched studies might have shown differences related to variations in study populations or design rather than the variable of interest.

All studies were judged to be of poor quality, based on reasons including: use of convenience data (i.e., opportunistic cohorts of men having prostate biopsy); potential biases (e.g., verification, selection, spectrum); incomplete or unclear study protocol (e.g., inclusion criteria, missing key variables); limited analyses and no matched analyses (or raw data from which to conduct matched analysis); and/or lack of blinding or reporting on blinding. In addition, four of the 24 studies addressing KQ 1 and KQ 2 were funded by GenProbe and a third of the 24 (N=8) indicated conflicts of interest for investigators (Table 1). All of these studies focused on determining the diagnostic accuracy of PCA3 testing using biopsy results as the reference or gold standard. No studies were identified that reported on intermediate outcomes other than diagnostic accuracy, or long-term clinical outcomes.

Evaluation of PCA3 and Other Comparators To Identify KQs 1 and 2 Intermediate and Long-Term Outcomes

Comparator: Total Serum PSA

Study design was a crucial criterion for this comparison, because tPSA measurements were integral to decisionmaking regarding uptake of prostate biopsy after the finding of an initial tPSA elevation through prostate cancer screening. Men were likely offered biopsy based on the extent of tPSA elevations, suspicious findings on a digital rectal exam (DRE), a combination of the two or, less commonly, other risk factors such as family history or race. This led to only a subset of initially identified men having the "gold standard" test (biopsy) that defines one of the outcomes of interest – diagnostic accuracy. This association of test result with uptake of the diagnostic test has been labeled verification bias.

As noted, verification bias would have occurred in this setting because men with higher tPSA elevation were more likely to undergo biopsy compared with men with lower levels. Thus, test sensitivity would have been overestimated (as a higher proportion of cancers with lesser

elevations would not have been identified by biopsy). This bias would have underestimated specificity (or overestimated the false positive rate), because the larger number of men without cancer and negative biopsy results were not identified via biopsy. See Appendix J for a more complete description of verification bias, an example relevant to prostate cancer and tPSA elevations, a review of directly relevant literature, and an evaluation of what will, or will not be compromised in this comparison. Appendix J also contains a more complete description of the modeling used to overcome the major impact of verification bias and a more extensive comparison of PCA3 and tPSA test performance characteristics.

These analyses indicated that the relative performance of tPSA elevations (sensitivity at a given specificity) was, at most, modestly influenced by verification bias, but the tPSA cutoff level at which this performance occurred cannot be directly observed. We employed a simple model to determine approximate tPSA cutoff levels in the presence of verification bias. This bias would be less likely to have been an issue for the other comparators (e.g., %fPSA, PSA density), but the extent of this bias is likely related to the correlation between that comparator and tPSA measurements. In addition, this correlation may be low because these comparators were not routinely used in all men with a tPSA/DRE positive result and may, therefore, not be strongly associated with biopsy uptake. A second known bias, sampling bias, was related to a subset of studies that limited their reporting to men in the "grey zone" of tPSA measurements and was variably defined in these studies as between 2.5 and 10 ng/mL,[91,95,106] 2.0 and 20 ng/mL[98] or 4 and 10 ng/mL.[105] These studies would have underestimated the performance of tPSA compared with studies that included all men with elevated results. We accounted for this bias by stratifying results, when possible.

Total PSA and the Intermediate Outcome of Diagnostic Accuracy

Key Points

The extent of tPSA elevations was compared with PCA3 scores to determine their diagnostic accuracy to predict prostate biopsy results (cancer/no cancer). Measures included in the analyses were the sensitivity, specificity (or the false-positive rate equal to 1-specificity), and positive and negative predictive values. As a reminder, only studies in which the performance estimates for both comparators were made in the same population were included in the five analyses listed below.

- **Area under the curve (AUC).** Twenty studies (Table 10) reported AUC estimates for tPSA and PCA3 in the same population and the difference of the two [AUC(PCA3) – AUC(tPSA)] was computed. Overall, 18 of the 20 studies found a positive difference. The two[90,102] studies finding tPSA elevations to have a greater AUC were among the smaller studies. Removing the four studies[31,91,95,106] that restricted recruitment to the tPSA "grey zone" resulted in an AUC difference of 0.0865 in the remaining 16 studies (Table 10).
- **Reported median, interquartile range, range and estimated logarithmic means/standard deviations (SD).** Eight studies (Table 11) provided sufficient data for analysis, and none of these directly reported a logarithmic SD (most, if not all studies examining the distribution found both PCA3 and tPSA to be highly right skewed). The logs SDs were estimated from the ranges or inter-quartile ranges. The differences, reported as z-scores, indicated that one study[102] (the smallest) found tPSA to be slightly

better than PCA3 at separating populations of positive and negative prostate biopsies, while the remaining seven others found a larger difference in favor of PCA3.

- **Performance at a PCA3 cutoff score of 35.** Nine studies (Table 12) reported the sensitivity and specificity of PCA3 at this cutoff. We computed the difference in sensitivity (PCA3 – tPSA) when tPSA was held at the PCA3-related specificity. Eight of the nine studies reported a positive difference (median 16.3 percent, range -9.5 to 35 percent) favoring PCA3 (Table 12).
- **ROC curves - sensitivity/specificity.** Fourteen studies and 16 datasets (Table 13) provided a ROC curve, or data representing a ROC curve, for both markers. At a specificity of 50 percent, the difference in corresponding specificities (PCA3 – tPSA) was zero or positive for all included studies except the one performed in a majority black population.[90] Removing the four studies that restricted recruitment to the tPSA "grey zone"[31,91,95,106] and the one study performed in a majority black population[90], the difference in sensitivity favored PCA3 by 20 percent (range 0 to 39 percent).
- **Regression analysis.** Only one study provided sufficient data to apply the respective regression coefficients to create a relative odds ratio (OR) between the 25th and 75th centiles of the two distributions.[31] A second study[95] reported all but the inter-quartile range, and that was estimated from the first study so that both datasets could be evaluated. In both studies, the ratio of the ORs (PCA3 / tPSA) was greater than 1 (1.38 and 1.97). These two studies[31,95] both restricted recruitment to the tPSA "grey zone," so the results were likely to overestimate the relative superiority of PCA3 by underestimating tPSA performance.

Interpretation

The results of analyzing the literature regarding the matched analyses of PCA3 score versus extent of tPSA elevations was summarized in Table 8. A more complete description of how these data were computed has been provided in Appendix J. Table 8 compares the diagnostic accuracy of PCA3 scores and tPSA elevations to independently identify men who would have a positive biopsy (prostate cancer). In Table 8A, the false-positive rate (1-specificity) was held constant, while in Table 8B, the sensitivity (detection rate) was held constant. This display was chosen because an undetected cancer was not considered equivalent to a falsely positive prostate biopsy and, therefore, comparing a loss in sensitivity with a gain in specificity was difficult. The last column shows the difference between the two estimates (PCA3 – tPSA). When comparing the sensitivities (Table 8A), this column contains the improvement in prostate cancer detection. When comparing the false-positive rates, it contains the reduction in biopsies performed on men without prostate cancer.

For example, assume that one would like to set test sensitivities to 85 percent. Using the row with 85 percent sensitivity (shaded row, Table 8B), only 59 percent of men without cancer would be subject to biopsy with PCA3 testing (cutoff score of about 17). Using tPSA elevations, 79 percent of those men without cancer would be biopsied (cutoff of 1.9 ng/mL). This means that using PCA3 instead of tPSA elevations, the same proportion of cancers might be detectable while performing 20 percentage points fewer biopsies. Additional tables at fixed PCA3 and tPSA cutoff levels, individual risks and positive and negative predictive values at several difference prostate cancer rates can be found in Appendix J (Tables J1 through J4).

Table 8. Modeled comparison of PCA3 and tPSA elevations to identify men with prostate cancer, with either the false positive rate (1 – specificity) or the sensitivity held constant

A) False Positive Rate (FPR) Held Constant					
FPR (1 -Specificity)	**PCA3 Cutoff (Score)**	**PCA3 Sensitivity**	**tPSA Cutoff (ng/mL)**	**tPSA Sensitivity**	**Improvement in Prostate Cancer Detection**
80%	9.3	95.1%	1.8	86.8%	8.3%
70%	12.6	91.0%	2.5	78.1%	12.9%
60%	16.4	88.7%	3.2	72.2%	16.5%
50%	21.0	79.1%	4.1	60.5%	18.6%
40%	26.8	71.2%	5.2	50.8%	20.4%
30%	34.9	61.2%	6.8	39.8%	21.4%
20%	47.4	48.8%	9.3	28.0%	20.8%
B) Sensitivity Held Constant					
Sensitivity	**PCA3 Cutoff (Score)**	**PCA3 FPR (1-Specificity)**	**tPSA Cutoff (ng/mL)**	**tPSA FPR (1-Specificity)**	**Reduction in Biopsies**
95%	9.3	80.0%	1.1	89.8%	9.8%
90%	13.3	68.2%	1.5	85.1%	16.9%
85%	16.9	58.9%	1.9	78.7%	19.8%
80%	20.4	50.2%	2.3	72.5%	22.3%
70%	27.7	38.7%	3.2	60.1%	21.4%
60%	36.0	28.9%	4.1	50.0%	21.1%
50%	46.0	20.9%	5.3	39.5%	18.6%

DR = proportion of biopsy positive men with a PCA3 score or tPSA elevation at or above the cutoff level; FPR = proportion of biopsy negative men with a PCA3 score or tPSA elevation at or above the cutoff level; tPSA = total prostate specific antigen

Characteristics of Studies Reporting Data Used in Five Analyses for KQs 1 and 2 Combined

Twenty two studies addressing KQ 1 and KQ 2 reported PCA3 and tPSA comparisons that could be used in one or more of the five analyses of matched studies (Table 9). Of interest are the five studies that used an upper cutoff for tPSA elevations to define a "grey zone."[31,91,95,98,106] When the tPSA range was truncated, it would reduce the effectiveness of the marker to predict biopsy outcome. In general, this was confirmed in our analyses, and, for that reason, the results may be stratified by this characteristic, the entries shaded in tables, and the observations noted in Figures. All study quality ratings were poor.

PCA3 and tPSA: Area Under the Curve

Twenty studies and 22 datasets reported the diagnostic performance of PCA3 and extent of tPSA elevation among men with initially screen positive test results (elevated tPSA with or without positive DRE) to discriminate between positive and negative biopsy test results. These studies and related information are shown in Table 10. Five studies[65,91,92,106,111] in which all individuals already had one or more negative biopsies were among the included studies, in order to strengthen the analysis of PCA3 and tPSA elevations. The studies are ordered by effect size, the difference between the matched AUC estimates of PCA3 minus tPSA (positive numbers indicate PCA3 performed better, negative numbers indicate tPSA performed better).

All but two studies[90,102] found the matched AUC point estimate for PCA3 higher than that for tPSA. Those two studies were among the five smallest reported, with 62 and 105 enrollees, respectively. The largest effect size was reported by the smallest study of all[109], reporting matched results for only 32 men. The median AUC difference for all studies was 0.1055 (range -0.1389 to 0.2150). Only eight[31,65,95,99,101-103,108] of the 20 studies reported the matched p-values comparing the two AUCs. Using these as a guide, at least the 10 studies from row 11 (Mearini[101]) to the end of the table are likely to have been statistically significant. No study reported a statistically significant lower performance for PCA3. The study reporting a difference of -0.139 did not report a p-value, but did provide the respective 95%confidence intervals (CI) on the PCA3 and tPSA AUC estimates. These overlapped, indicating the differences were not likely to be significant (0.705, 95% CI: 0.599 to 0.812; and 0.844, 95% CI: 0.765 to 0.910, respectively).

Table 9. Summary results for five analytic comparisons[a] of PCA3 versus tPSA in matched populations of men having prostate biopsies

Author Study[b], Year	N	Initial Biopsy	AUC	Mean/SD	PCA3>35	Sens/Spec	Reg
Adam[90], 2011	105	82%	-0.1389	-	3.1%	1%	-
Nyberg[102], 2010	62	55%	-0.0543	-0.22	-23.3%	0%	-
Bolitto[93] (initial), 2012	735	100%	0.0010	-	-	0%	-
Aubin[91], 2010	1,072	0%	0.0310	-	12.9%	7%	-
Roobol[108], 2010	721	71%	0.0540	-	-	-	-
Rigau[107], 2010	215	74%	0.0580	-	-	11%	-
Ankerst[65], 2008	443	0%	0.0580	0.31	-	11%	-
Hessels[100], 2010	336	NR	0.0700	0.50	-	-	-
Wu[111], 2012	103	0%	0.0700	-	6.0%	28%	-
Ouyang[104],2009	92	NR	0.0800	-	-	-	-
Bolitto[93] (composite), 2012	1,246	59%	0.0930	0.47	9.5%	-	-
Goode[99] (repeat), 2012	169	0%	0.1050		-	26%	-
Mearini[101], 2009	96	NR	0.1180	-	-	-	-
Ochiai[103], 2011	105	81%	0.1264	0.98	19.8%	7%	-
Ploussard[106], 2010	301	0%	0.1350	-	-	21%	-
Auprich[92], 2011	127	0%	0.1350	1.05	-	-	-
Deras[96], 2008	557	51%	0.1390	-	21.4%	20%	-
Cao[94], 2011	131	NR	0.1730	-	-	-	-
de la Taille[95], 2011	516	100%	0.1840	-	31.0%	26%	1.38
Perdona[31], 2011	218	61%	0.1920	0.38	-	25%	1.97

Table 9. Summary results for five analytic comparisons[a] of PCA3 versus tPSA in matched populations of men having prostate biopsies (continued)

Author Study[b], Year	N	Initial Biopsy	AUC	Mean/SD	PCA3>35	Sens/Spec	Reg
Schilling[109], 2010	32	86%	0.2000	-	-	39%	-
Goode[99] (composite), 2012	456	63%	0.2150	-	35.0%	-	-
Goode[99] (initial), 2012	287	100%	0.2200	-	-	24%	-
Bolitto[93] (repeat), 2012	511	0%	0.2310	-	-	29%	-
Wang[110], 2009	187	73%	-	-	28.4%	-	-
Ferro[98], 2012	151	100%	-	0.44	-	-	-

N = number; NR = not reported; Sens = sensitivity; Spec = specificity; tPSA = total prostate specific antigen
[a]AUC = area under the curve for PCA3 minus the AUC of tPSA; Mean/DS = difference in separation between PCA3 scores and tPSA results, when expressed as z-scores; PCA3>35 = difference of the PCA3 minus the tPSA sensitivities at the specificity found for a PCA3 cutoff of 35; Sens/Spec = difference between PCA3 and tPSA sensitivity at a specificity of 50%; Reg = relative change in the PCA3 ORs (between the 25th and 75th centiles) and the corresponding tPSA ORs. The corresponding full analyses resulting in these summaries can be found on the following pages.
[b]Order based on AUC difference; shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients. Rows 4, 15, 19, 20, and 26 are shaded.

None of the studies reported a confidence interval or standard deviation for the matched difference of the two AUCs. Although the AUCs for PCA3 and tPSA ranged widely (indicating relatively high heterogeneity), the variability of the differences seemed more consistent. This may be due to the requirement that only paired estimates of the AUCs be included in this analysis.

Four studies[31,91,95,106] enrolled only men with tPSA levels less than 10 ng/mL, essentially limiting their population to the so-called "grey zone." In general, only one to two percent of biopsy negative men had tPSA levels over 10, while about 20 percent of biopsy positive men were in this range.[120,121] Removing this subset from the overall population of men with positive tPSA/DRE was likely to have the effect of reducing the ability of tPSA to predict positive prostate biopsies. Thus, one would have expected these studies to show greater differences in favor of PCA3. The four studies focusing on the "grey zone" are highlighted in grey in Table 10. All but one[91] was near the bottom of the table, indicating that they did, in fact, find greater differences. The median difference in AUC in the "grey zone" studies was 0.1595. If these four studies were removed, the AUC difference was reduced to 0.0865. Although not formally computed, the heterogeneity would also be expected to be reduced. One could argue that these four "grey zone" studies should have been excluded, as they did not, technically, satisfy fully the inclusion criteria. However, they were included for two reasons. First, the performance in this subset had clinical implications. For example, some may argue that clinicians could intervene based solely on a very elevated tPSA, but use additional markers to evaluate the remaining "grey zone" patients. This assumes that very elevated tPSA results are, by themselves, sufficiently informative for decisionmaking, and performance would not benefit from adding a second useful and independent marker like PCA3. Should PCA3 come into routine practice, it is not clear that use only in the "grey zone" would be an effective approach. Second, the "grey zone" stratification identified a source of heterogeneity and helped demonstrate the validity of these analyses.

An estimate of the potential for publication bias for this analysis could be generated under the assumption that the standard error of the AUC difference was proportional to the reciprocal of the square root of the number of enrolled men for each study. Figure 7 shows a plot of the

computed AUC difference (x-axis) versus its estimated precision (i.e., the reciprocal of the square root of the sample size [y-axis]). The solid vertical line shows the median difference of 0.1055 while the dashed vertical line at 0.000 shows where the AUC would be equivalent. As predicted, the data were found to fit a symmetric "inverted funnel," suggesting that at least some of the variability was due to the small samples sizes for several of the studies. The results seemed far more consistent for the 11 largest studies[31,65,91,93,95,96,99,100,106-108] that provided matched results for 200 or more men.

Table 10. Comparing PCA3 scores and tPSA elevations in matched studies via AUC analysis

Table 10a. Results from 20 matched studies reporting results for all study subjects[a]

Author Study[b], Year	Number	Initial Bx	PCA3 AUC	tPSA AUC	Difference	P-value[c]
Adam[90], 2011	105	82%	0.7054	0.8443	-0.1389	-
Nyberg[102], 2010	62	55%	0.7418	0.7961	-0.0543	0.07
Aubin[91], 2010	1,072	0%	0.6430	0.6120	0.0310	
Roobol[108], 2010	721	71%	0.6350	0.5810	0.0540	0.14
Ankerst[65], 2008	443	0%	0.6650	0.6070	0.0580	>0.05
Rigau[107], 2010	215	74%	0.6600	0.6020	0.0580	-
Hessels[100], 2010	336	-	0.7200	0.6500	0.0700	-
Wu[111], 2012	103	0%	0.6400	0.5700	0.0700	
Ouyang[104], 2009	92	-	0.6700	0.5900	0.0800	-
Bollito[93], 2012	1246	59%	0.6780	0.5850	0.0930	
Mearini[101], 2009	96	-	0.8140	0.6960	0.1180	<0.05
Ochiai[103], 2011	105	81%	0.8507	0.7243	0.1264	0.025
Ploussard[106], 2010	301	0%	0.6880	0.5530	0.1350	-
Auprich[92],2011	127	0%	0.7030	0.5680	0.1350	-
Deras[96], 2008	557	50%	0.6860	0.5470	0.1390	-
Cao[94], 2011	131	-	0.7390	0.5660	0.1730	-
de la Taille[95], 2011	516	100%	0.7610	0.5770	0.1840	<0.001
Perdona[31], 2011	218	61%	0.8280	0.6360	0.1920	<0.001
Schilling[109], 2010	32	56%	0.8100	0.6100	0.2000	-
Goode[99], 2012	456	63%	0.7260	0.5110	0.2150	<0.001
Median Difference (all data, N=20)	**6,934**				**0.1055**	
Median Difference (only "grey zone" studies; N=4)	**2,107**				**0.1595**	
Median Difference (excluding "grey zone" studies; N=16)	**4,827**				**0.0865**	

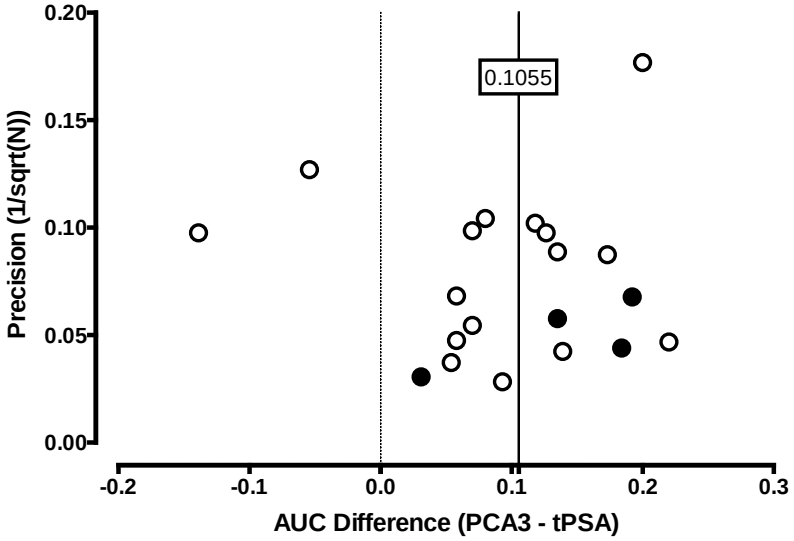
AUC = area under the curve; Bx = biopsy; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
A dash ('-') = no value reported.
[a]Comparisons of difference for Figure 5A were between all reported subjects, regardless of repeat/initial biopsy status.
[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (2.5-10 ng/mL) when enrolling patients. In Table 10a, rows 3, 13, 17, and 18 are shaded.
[c]Reported p-value for the comparison of the two AUCs computed among the same set of men.

Table 10b. Results from three matched studies reporting results stratified by biopsy status[a]

Author Study[b], Year	Number	Initial Bx	PCA3 AUC	tPSA AUC	Difference	P-value[c]
Bollito[93], 2012	511	0%	0.7840	0.5530	0.2310	-
Bollito[93], 2012	735	100%	0.6140	0.6130	0.0010	-
Goode[99], 2012	169	0%	0.6050	0.5000	0.1050	-
Goode[99], 2012	287	100%	0.7720	0.5520	0.2200	-
Perdona[31], 2011	85	0%	0.7480	0.5960	0.1520	-
Perdona[31], 2011	133	100%	0.8730	0.6600	0.2130	-

AUC = area under the curve; Bx = biopsy; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
A dash ('-') = no value reported.
[a]Comparisons of difference for Figure 5B replaced composite proportions for four studies with results from initial and repeat biopsy subgroups
[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (2.5-10 ng/mL) when enrolling patients. In Table 10b, rows 5 and 6 are shaded.
[c]Reported p-value for the comparison of the two AUCs computed among the same set of men.

Figure 7. Examining the relationship between effect size (AUC difference) and an estimate of precision

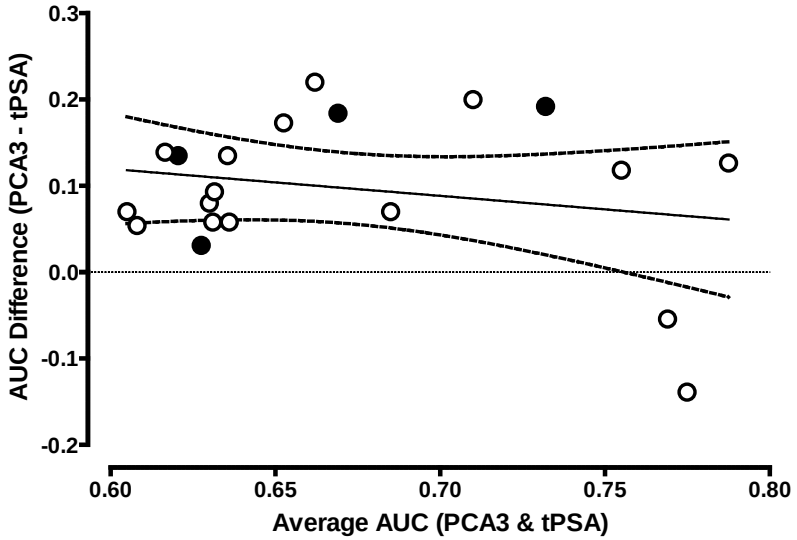


AUC = area under the curve; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
Note: In this analysis, precision is estimated by the reciprocal of the square root of N (the number of study subjects). The vertical dashed line indicates the points at which the two tests (PCA3 and tPSA) perform equally. The filled circles indicate studies focusing on the "grey zone" of tPSA.

Figure 8a explores the relationship of the AUC difference, this time comparing the results against the average AUC (average of PCA3 and tPSA AUCs). The actual AUC for the markers may be indicative of extraneous factors (e.g., tPSA cutoff level, age of enrollees) that may vary among the 20 studies included in this analysis. Of interest, the two studies[90,102] reporting the negative AUC differences were two of the three highest average AUCs. However, regression analysis showed no significant relationship between average AUC and the AUC difference (slope not significant; p=0.72). The median average AUC was 0.6605 (range 0.5525 to 0.7875). Figure 8b displays the tPSA AUC on the x-axis versus the PCA3 AUC on the y-axis for the same studies shown in Figure 8a. The dashed "line of identity" indicates all values where the tPSA and PCA3 AUCs would be equal. On average, the observations fall above the line, showing that the PCA3 AUC is higher than the tPSA AUC within a given study. As expected, the four studies in

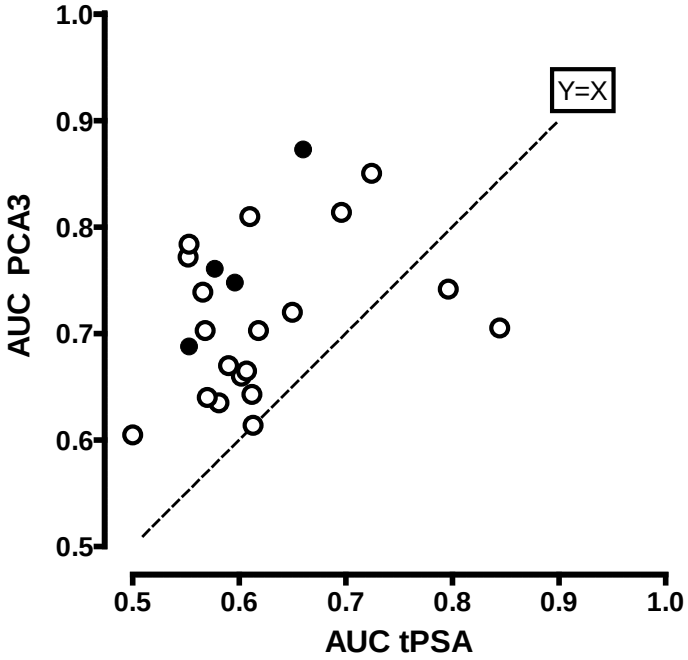
the “grey zone” of tPSA (filled circles) tend to have higher differences that fall farther from the line of identity.

Figure 8a. The relationship between AUC difference and average AUC for PCA3 and tPSA



AUC = area under the curve; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
Note: The solid line indicates the results of linear regression (AUC_Difference = -0.3120 * Avg_AUC + 0.3068). The dashed lines show the 95% prediction limits. The filled circles indicate studies focusing on the “grey zone” of tPSA .

Figure 8b. Scatterplot showing the reported tPSA AUC versus the reported PCA3 AUC
AUC = area under the curve; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen



Note: The dashed line shows the points at which the two AUCs would be equivalent. The majority of observations are above the line, indicating the PCA3 AUCs are, on average, higher. The filled circles indicate studies focusing on the “grey zone” of tPSA.

Eighteen studies identified the methodology used for PCA3 testing (Table 4). Two used Aptima reagents (GenProbe),[100,103] 10 specified using the Progensa® kit,[31,90,92,93,95,96,99,102,106,108] and two reported only using a "GenProbe" test.[91,109] Among the remaining six studies, four used a quantitative RT-PCR method,[94,101,104,107] The remaining two studies[64,106] did not specify the method, but disclosures suggested that both were likely using the current GenProbe assay (Progensa).[65,111] AUC differences (PCA3-tPSA) for fourteen studies using GenProbe reagents were compared with the proportion of men having an initial biopsy (figure not shown). This regression analysis showed no significant relationship between the AUC difference and initial biopsy status (slope -0.0472; p=0.52). The slope indicates that over the range of AUCs shown in Figure 8a (0.6 to 0.8), the difference in AUC would be -0.0094 or about a 1 percent lower for all initial biopsied patients compared with all repeat biopsied patients. The same analysis for six studies using other assays also showed no significant relationship (slope 0.0854; p=0.27), but the slope indicated a 1.7 percent higher AUC for the same comparison. Therefore, the analysis does not provide statistically significant evidence that assay methodology is an important consideration.

Among the six of 20 studies that reported the racial/ethnic distribution in the study population (Table 3),[65,90,94,96,103,104] one was in an Asian (Japanese) population[103] and this group's AUC difference of 0.126 was slightly higher than the summary estimate of 0.1055 Another Asian study[94] (China) had an AUC difference estimate of 0.173. Only one study performed in South Africa reported on a population composed of a majority of black men (68.6 percent, Table 3),[90] and the group's AUC difference of -0.139 was the lowest observed in all studies (Table 10). The four North American studies reporting a small black population (5.3[96,110] and 2 percent[65,104]) had AUC differences near the consensus estimate.

Each reviewed study was assigned a QUADAS quality score of good, fair, or poor. Among the 20 included studies in the AUC differences computation (Table 10), all were rated poor. Only one of the studies was blinded in both directions (i.e., laboratory blinded to outcome, and clinicians blinded to laboratory results), and only two were blinded in a single direction (one in each direction).

PCA3 and tPSA: Reported Medians and Standard Deviations

Eight studies (Table 11) reported some information concerning the distributions of PCA3 and tPSA levels among men with screen positive test results (elevated tPSA with or without positive DRE) who subsequently had positive or negative biopsy results. Results from these studies and related information are also shown in Table 11. The distributions of both markers are highly right skewed and have been shown to be reasonably Gaussian after a logarithmic transformation. For this reason, we chose to include for analysis only those studies in which the median or logarithmic mean could be determined along with the logarithmic standard deviation. In some instances the standard deviation was estimated using reported centiles (e.g., inter-quartile range). If a study only reported the range, the standard deviation was computed assuming the range represented 6 standard deviations.[122] For each study, the difference in marker levels in those with positive or negative biopsies was expressed as a z-score, using a study-specific pooled standard deviation.

It was possible to obtain a median and pooled log standard deviation for both markers using data from eight studies (Table 11). Studies were sorted by the difference in z-scores. The two studies that truncated results above 10 ng/mL[31] or 20 ng/mL[98] are shown in grey. One study[65] incorrectly reported the median PCA3 score in men with a negative biopsy; the corrected value

of 19.4 is shown. Two additional studies[91,95] had partial data, and were also summarized at the bottom of Table 11 to allow for comparison of median levels only. This analysis would have been more robust had the authors actually reported the median and logarithmic standard deviations for their populations, provided raw data (in the form of a scatterplot) or fitted the data to some other distribution.

Figure 9 shows the overlapping distributions from the eight studies shown in Table 11. The overlapping curves were drawn based on the log Gaussian parameters described there. The individual figures show each set of paired distributions. Given the fact that only eight studies were analyzed, it was not possible to stratify results by race, region or test methodology. Note that the very tight distributions for tPSA can be seen for the two "grey zone" studies.[31,98]

PCA3 and tPSA: Performance at a PCA3 Cutoff Score of 35

Nine studies (Table 12) reported the sensitivity and specificity of PCA3 score at a cutoff of 35 among men with positive initial screening results (elevated tPSA with or without positive DRE) who subsequently had positive or negative biopsy results. Table 12 shows the sensitivity and false positive rates (1 – specificity) for PCA3 with the corresponding sensitivity of tPSA at the same specificity found for the PCA3 cutoff level. The table was sorted by effect size. The difference in the two sensitivities (with the specificity held constant) provided a comparison of the ability to distinguish prostate cancer between the two markers. In some instances, the tPSA results were estimated from a published ROC curve. In one study, the specificity/1-specificity was incorrectly reported, as evidenced by the additive inverse found on the accompanying ROC curve. In another study[102], the reported sensitivity/specificity did not match the corresponding ROC curve, and the reason for the discrepancy could not be identified. Those data were excluded from analysis. The most appropriate analysis that compared two tests on the same population was to use a matched analysis of the 2x2 table. However, all nine studies reported only independent evaluations of each marker.

Among the nine studies (Table 12), the PCA3 score cutoff level of 35 was associated with false-positive rates (1-specificity) ranging between 20 and 50 percent, with corresponding sensitivities (detection rates) between 38 and 77 percent. For each study, the corresponding sensitivity for tPSA (at the same false positive rate) was subtracted from the PCA3 sensitivity. For one study[90], the difference was negative, while the eight remaining studies showed PCA3 having higher sensitivities with increases ranging from 3 to 35 percent. The median increase in sensitivity was 16.3 percent.

Given that only nine studies were analyzed, it was not possible to stratify results by race, region or test methodology. Of interest, however, is that the one study[90] that found PCA3 to be least useful was performed in a largely black population and was quite small (45 positive biopsies), leading to a wide confidence interval on the sensitivity estimate.

Table 11. Comparison of PCA3 and tPSA differences in central estimates in men with positive and negative prostate biopsy results, after accounting for study-specific variability in measurements

Author Study[a], Year	N	Median PCA3 Score: Positive Bx	Median PCA3 Score: Negative Bx	Median PCA3 Score: Pooled Log SD	Median PCA3 Score: Z_{PCA3}[b]	Median tPSA (ng/mL): Positive Bx	Median tPSA (ng/mL): Negative Bx	Median tPSA (ng/mL): Pooled Log SD	Median tPSA (ng/mL) Z_{PCA3}[b]	$Z_{PCA3} - Z_{tPSA}$
Nyberg[102], 2010	62	49.0	22.0	0.4150	0.84	12.6	6.2	0.2918	1.06	-0.22
Ankerst[65], 2008	443	34.3	19.4	0.4480	0.55	8.2	6.7	0.3604	0.24	0.31
Perdona[31], 2011	218	72.0	22.0	0.4264	1.21	8.0	6.0	0.1514	0.83	0.38
Ferro[98], 2012	151	57.0	28.0	0.3469	0.89	7.9	6.8	0.1438	0.45	0.44
Bollito[93], 2012	1,246	63.0	35.0	0.4530	0.56	7.4	6.3	0.6873	0.09	0.47
Hessels[100], 2010	336	50.0	18.0	0.5492	0.81	8.3	5.9	0.4836	0.31	0.50
Ochiai[103], 2011	105	59.5	14.2	0.3489	1.78	10.7	6.3	0.2875	0.80	0.98
Auprich[92], 2011	127	75.0	35.0	0.2929	1.13	8.8	8.1	0.4605	0.08	1.05
Number / Medians	**2,687**	**58.3**	**22.0**			**6.3**	**8.3**			
Aubin[91], 2010 [c]	1,072	33.8	16.7	-	-	-	-	-	-	-
de la Taille[95], 2011 [c]	516	50.0	18.0	-	-	5.8	5.2	-	-	-

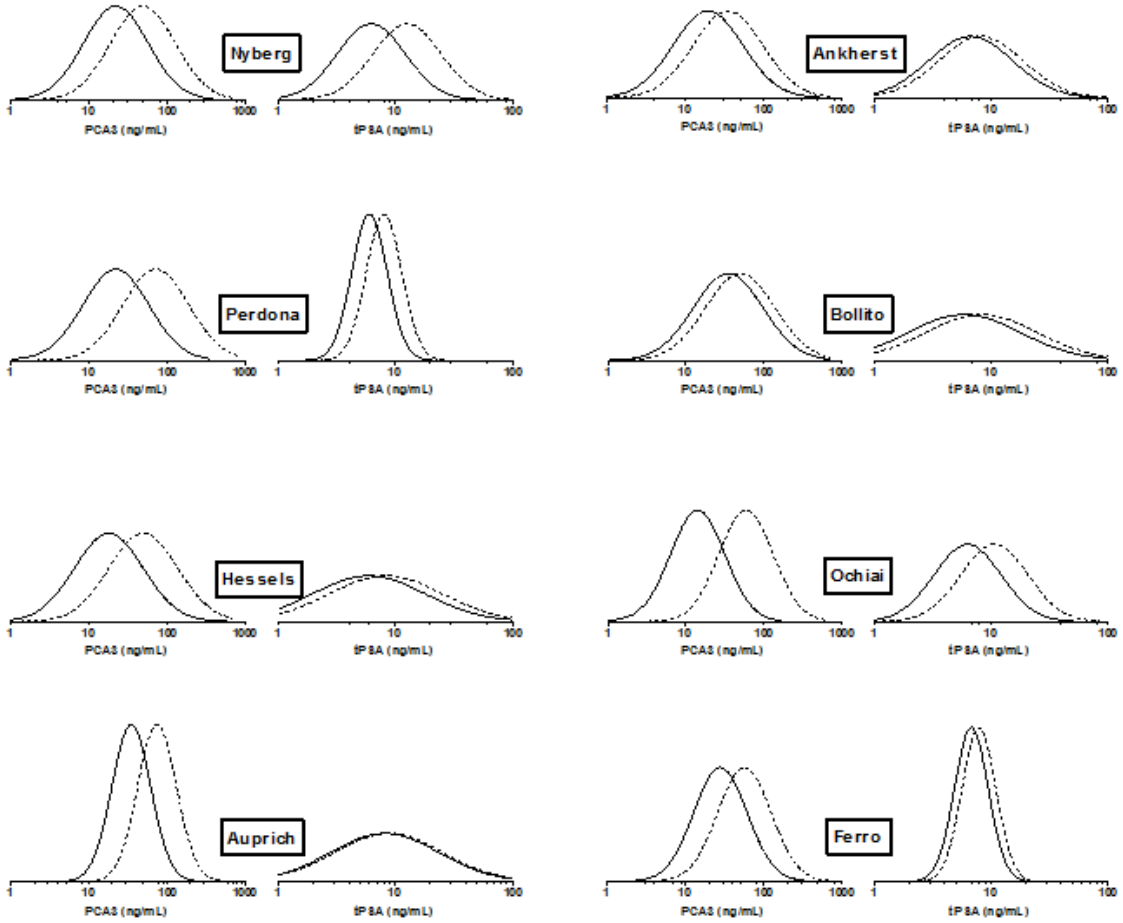
Bx = prostate biopsy; N = number; SD = standard deviation; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen

[a]Ordered by $Z_{PCA3} - Z_{tPSA}$ result; shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients. Rows 3, 4, 10, and 11 are shaded.

[b]Z score = (log (Positive Bx median) – log (Negative Bx median)) / pooled log SD within each study.

[c]The last two studies[91,95] reported partial information for comparison of biopsy positive and negative median values, but did not provide sufficient data to compute Z scores.

Figure 9. Distributions of PCA3 and tPSA in men with positive and negative prostate biopsies from four studies that reported such information in the same cohort



PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
Note: The logarithmic means and pooled standard deviations were obtained from Table 11. The solid line indicates the distribution for those with a negative biopsy, while the dashed line represents the distribution in those with a positive biopsy. For each study (indicated by the first author), these distributions were obtained in the same group of men. The figures are sorted by the relative performance of PCA3 over tPSA (from left to right, top to bottom).

As a way of estimating whether the nine studies were reasonably consistent in their estimates of sensitivity and specificity, a summary analysis was performed for PCA3 (Figure 10). There was high and significant heterogeneity (I^2=100 percent. p<0.001). This can be seen in the figure and the table with the two studies[90,93] having much higher false positive rates, but only modestly higher sensitivities. Thus, a better summary of the data is the fitted ROC curve shown in Figure 10 [Spearman correlation between the logit (sensitivity) and logit (1-specificity) = 0.76, p=0.01]. This presentation is not subject to the usual strong bias introduced by the tPSA upper cutoff of 10 ng/mL used in two of the studies,[91,95] as PCA3 is essentially independent of tPSA measurements (Table 15), and Figure 10 focuses only on the PCA3 results.

Table 12. Differences in PCA3 and tPSA sensitivities at the fixed specificity associated with the commonly used PCA3 score cutoff of 35[a]

Study Author[b,c], Year	N	Initial Bx	Positive Bx	PCA3 1-Spec (%)	PCA3 Sens (%)	tPSA Sens (%)[a]	Difference (%)
Adam[90], 2011	105	82%	42.9%	50.0	77.7	87.2	-9.5
Wu[111], 2012	103	0%	36.0%	23.0	38.0	32.0	6.0
Bollito[93], 2012	1,246	59%	25.9%	49.1	72.0	62.5	9.5
Aubin[91], 2010	1,072	0%	17.7%	21.4	48.4	35.5	12.9
Ochiai[103], 2011	105	81%	36.0%	25.4	74.3	54.5	19.8
Deras[96],2008	557	50%	36.0%	26.1	53.9	32.5	21.4
Wang[110], 2009	187	73%	46.5%	20.0	52.9	24.5	28.4
de la Taille[95], 2011	516	100%	40.0%	24.0	64.0	33.0	31.0
Goode[99], 2012	456	63%	19.3%	25.0	62.0	27.0	35.0
Median Difference	**4,409**						**16.3**

Bx = prostate biopsy; Diff = (PCA3 % Sens – tPSA % Sens); N = number; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen; Spec = specificity (1-specificity=false positive rate); Sens = sensitivity (detection rate)

[a]Sensitivity for tPSA elevation at the same 1-specificity (false positive rate) found for a PCA3 score at a cutoff of 35 in that study.

[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients. Rows 4 and 8 are shaded.

[c]Note that a small study by Nyberg et al[102] was deleted from this table because the reported PCA3 sensitivity and 1-specificity were not consistent with the reported PCA3 ROC curve.

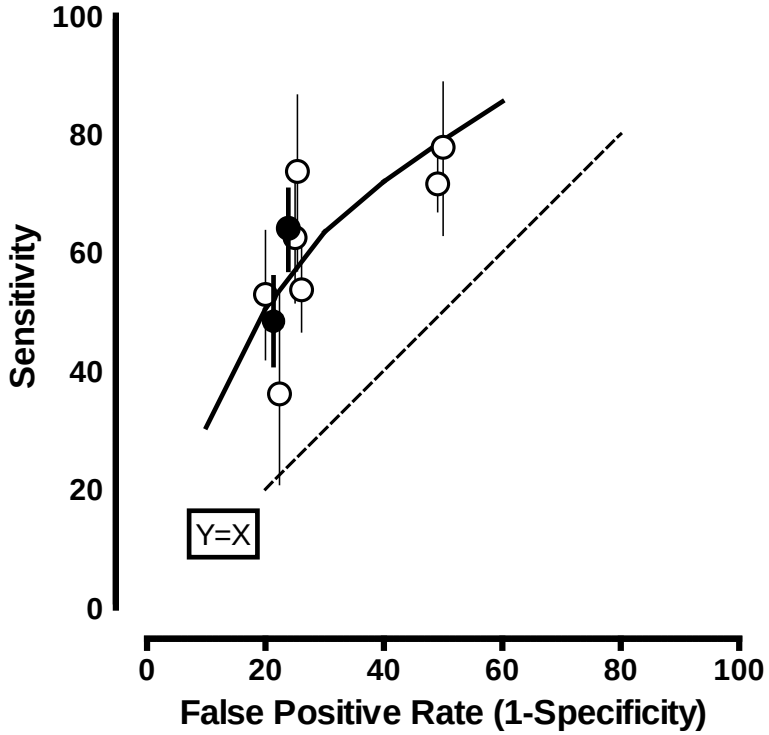
Note: The number of true positives = N * Positive * Sens (*e.g.,* for Adam, TP=105*42.9% * 77.7% = 35); FN=N*Positive-TP; TN=N*(100-Positive)*Spec; FP=N*(100-Positive)-TN.

PCA3 and tPSA: ROC Curves-Sensitivity/Specificity

Fourteen studies (Table 13) provided ROC curves for both PCA3 scores and tPSA elevations among men with positive initial screening test results (elevated tPSA with or without positive DRE) who subsequently had positive or negative biopsy results. Two of these studies[93,99] reported ROC curves separately for initial and repeat biopsies and, therefore, there are 16 rows/datasets. The performance of PCA3 and tPSA testing are presented in Table 13, sorted from smallest to largest number of enrolled men. For each study, the sensitivities of each marker at preselected false positive (1-specificity) rates were estimated from published ROC curves. These values were recorded to the nearest percent (e.g., sensitivity of 55 percent). The table entries showing test performance are the PCA3 sensitivity, followed, in parentheses, by the incremental increase, or decrease, of tPSA sensitivity. For example, at a false-positive rate (1-specificity) of 20 percent, the first study found a PCA3 sensitivity of 57 percent, which was 19 percent higher than tPSA sensitivity, (i.e., 57 - 19 = 38 percent). Negative numbers indicated that tPSA is performing better; positive numbers indicated PCA3 is performing better.

The last three lines in Table 13 are the median results ignoring matching. That is, the median PCA3 sensitivity is provided along with the median difference computed separately. The first of the three lines summarizes all 13 studies. The next summarizes the four tPSA "grey zone" studies, while the last summarizes the nine remaining studies (non-shaded rows) after the one study[90] performed in a mainly black population was removed.

Figure 10. The performance of PCA3 score to identify subsequent positive prostate cancer biopsies from nine studies



Note: The sensitivity (y-axis) versus the false positive rate (1-specificity, x-axis) of PCA3 testing is shown for the nine studies. Filled circles indicate the study that focus on the "grey zone" of tPSA (2.5-10 ng/mL) when enrolling patients. The thin solid lines indicate the 95% confidence intervals of the sensitivity estimates. The solid line is the fitted ROC curve to the provided data. For reference, the dashed line indicates where the sensitivity and 1 - specificity are equal, indicating a useless test.

Table 13. Differences in PCA3 and tPSA sensitivities (difference)[a] at PCA3 false positive (1-specificity) rates from 20% to 80%

Study Author[b]	N	% Initial Bx	20%	30%	40%	50%	60%	70%	80%
Schilling[109]	32	56	57 (19)	71 (29)	90 (45)	95 (39)	95 (17)	100 (6)	100 (6)
Nyberg[102]	62	55	50 (-11)	61 (-11)	73 (-4)	89 (0)	94 (-3)	94 (-6)	94 (-6)
Wu[111]	103	0	37 (13)	45 (5)	67 (27)	77 (35)	82 (29)	82 (9)	82 (-2)
Adam[90]	105	82	50 (-19)	68 (-7)	75 (0)	78 (-9)	78 (-13)	80 (-18)	90 (-10)
Ochiai[103]	105	81	74 (23)	87 (30)	92 (20)	97 (7)	98 (9)	100 (8)	100 (8)
Goode[99] (repeat)	167	0	34 (13)	43 (11)	56 (24)	73 (26)	79 (14)	85 (15)	90 (5)
Rigau[107]	215	74	32 (-2)	57 (15)	64 (14)	77 (11)	84 (11)	87 (9)	93 (10)
Perdona[31]	218	61	70 (24)	79 (21)	82 (21)	88 (25)	97 (20)	100 (17)	100 (12)
Goode[99] (initial)	289	100	60 (37)	77 (44)	82 (43)	82 (24)	88 (33)	95 (19)	98 (3)
Ploussard[106]	301	0	40 (12)	56 (21)	71 (25)	80 (21)	83 (13)	89 (11)	93 (11)
Ankerst[65]	443	0	38 (-1)	51 (3)	67 (9)	76 (11)	83 (9)	88 (7)	95 (9)
Bollito[93] (repeat)	509	0	69 (42)	76 (31)	81 (31)	86 (29)	91 (22)	91 (15)	93 (8)
de la Taille[95]	516	100	57 (27)	71 (29)	77 (27)	85 (26)	89 (21)	93 (17)	95 (7)
Deras[96], 2008	557	50	45 (21)	55 (22)	63 (19)	74 (20)	82 (25)	89 (13)	95 (5)
Bollito[93] (initial)	728	100	42 (2)	48 (-3)	57 (-1)	61 (0)	66 (-1)	77 (-2)	90 (1)
Aubin[91]	1,072	0	49 (13)	60 (18)	68 (11)	74 (7)	79 (5)	85 (2)	94 (3)
Medians (13 studies, 16 datasets)	**4,979**		**49 (13)**	**61 (19)**	**72 (21)**	**79 (20)**	**83 (15)**	**89 (9)**	**94 (5)**
Medians (4 'grey zone' studies)	**2,107**		**53 (19)**	**65 (21)**	**74 (23)**	**83 (23)**	**86 (17)**	**91 (14)**	**95 (9)**
Medians (9 studies, 11 datasets)[c]	**2,810**		**45 (13)**	**57 (15)**	**67 (20)**	**77 (20)**	**84 (17)**	**89 (9)**	**94 (5)**

Bx = prostate biopsy; N = number; tPSA = total prostate specific antigen

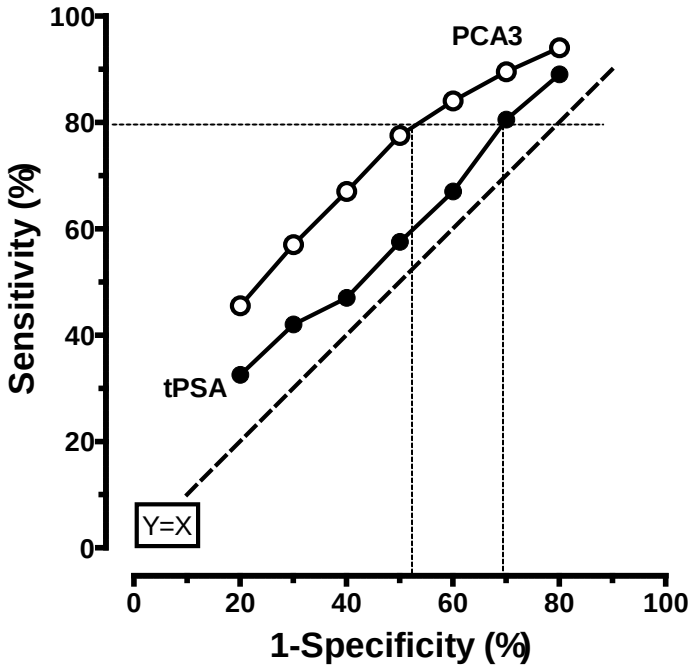
[a]Difference = (PCA3 sens – tPSA sens) when (1-specificity) is held constant at values ranging from 20% to 80%. For example, at a 1-specificity of 50%, the PCA3 sensitivity in the Schilling study is 95% and the tPSA sensitivity is 39% lower or 56%.

[b]Shaded rows indicate studies focusing on the "grey zone" of tPSA (e.g., 2.5-10 ng/mL) when enrolling patients, as well as the summary line where the results of those studies are summarized. Rows 8, 10, 13, 16, and 18 are shaded.

[c]After removal of Adam[90], a study focusing on a South African 69% black population; 11 estimates since 2 each for Goode & Bollito.

Figure 11 displays the summary ROC curves computed using the PCA3 median sensitivities and median differences provided in the last row of Table 13. This can be taken as a simple summary of performance for the two tests, under similar circumstances in a general population of men with elevated tPSA.

Figure 11. Summary ROC curves for PCA3 and tPSA based on studies applying the tests to the same population



Note: The open circles (solid line) indicates performance of PCA3 scores, while the filled circles (solid line) indicated tPSA performance. The thick dashed line indicated where the sensitivity equals 1-specificity, indicating a test with no predictive ability. Data are presented in Table 13.

PCA3 and tPSA: Regression Analysis

Two studies[31,95] reported sufficient results of regression analysis separately for PCA3 and for tPSA elevations in the same population of men to be included in these analyses. Both studies included in this analysis restricted tPSA levels to less than 10 ng/mL ("grey zone"). Each of the studies reported the odds ratio (OR) for each marker when that marker was assumed to be a continuous variable. That is, the antilog of the OR will be the regression coefficient per unit increase of the marker (e.g., increase of PCA3 score from 30 to 31). This makes comparison of PCA3 and tPSA difficult, as the range of results for the two markers differs. To account for this, the coefficients will be used to compute the ratio of the ORs at the 25th and 75th centiles for each marker. This is a measure of the change in odds over the inter-quartile range. This ratio of ORs for PCA3 will then be divided by the corresponding ratio for tPSA. Values greater than 1 indicates that PCA3 provided more discrimination than tPSA. This normalization also allows for comparisons between studies, where the coefficient is dependent on the range of tPSA values studied.

Only one of the included studies[31] provided the inter-quartile ranges for both markers. It was necessary to estimate those ranges for the second study.[95] For PCA3, this was done by extrapolating the log mean and SD from two centiles provided as part of the sensitivity/ specificity results. For tPSA, this was done by using the inter-quartile range from the first study[31] and adjusting for a minor difference in the mean values reported.

Two additional studies provide some further insight. One[91] showed similar coefficients for PCA3 and tPSA, but it was not possible to compute the ratio of the ORs for tPSA because no data were provided to estimate the 25th and 75th centiles. However, given that the inter-quartile range of PCA3 scores were generally larger than the corresponding range of tPSA results, these coefficients were likely to have shown an overall finding of PCA3 being more discriminatory. Another study[94] provided only the continuous OR estimates. The PCA3 OR was the highest reported among the four studies in Table 14, and the corresponding OR for tPSA was slightly under 1.0. This would have to be associated with PCA3 being more discriminatory, but estimates of effect size could not be provided because information about the distributions were not provided.

Five studies[31,91,94,95,103] provided some information on the independence of PCA3 and tPSA as markers for prostate biopsy status (Table 15). Two specific measures were sought. Thought to be most useful were the bivariate correlations (parametric or non-parametric) between the two markers for those with positive, and for those with negative, prostate biopsies. Alternatively, logistic regression coefficients (or the corresponding ORs) reported with, and without, adjustment for tPSA were evaluated. In many of the studies reporting logistic regression models, additional factors such as history and prostate volume were also included. If both PCA3 and tPSA coefficients remained essentially constant after adjusting for the other marker (and possibly additional markers), this was taken as evidence that the two markers together were more predictive than either alone (independent).

Three studies[31,94,95] reported information on correlation coefficients (Table 15). One reported the two correlation coefficients (non-parametric estimates),[94] two reported a single merged correlation (parametric)[95], and the third just reported that the correlations were "low" for both groups.[31] One potential problem with these estimates is that reliable correlation estimates for both PCA3 and tPSA would require a logarithm transformation prior to computing a parameter estimate such as the Pearson's correlation coefficient. None reported that the data were transformed. Overall, the two markers were not highly correlated in either of the groups of interest.

Five studies[31,91,94,95,103] provided information on coefficients for PCA3 and/or tPSA from univariate and multivariate logistic regression modeling (Table 15). Four of the five[31,91,95,103] found the PCA3 coefficients unchanged after accounting for tPSA (and often other variables as well). The remaining study[94] found a reduction in the coefficient, but it was still the most significant predictor. In addition, this study did not include tPSA in the multivariate model, as it was not statistically significant in the univariate logistic regression ($p=0.08$).

The results were less consistent for tPSA. Three studies[31,91,95] found tPSA essentially unchanged after accounting for PCA3. Interestingly, these three studies all restricted tPSA levels to under 10 ng/mL. This may reduce the correlation between the two markers, if PCA3 and tPSA are concordant when tPSA elevations are relatively high. A fourth study[103] did not report the coefficients but did report that the p-value was reduced from being highly significant ($p<0.001$) to no significance ($p=0.52$). The fifth[94] did not report results for tPSA after adjustment, as it only included variables found to be significant in univariate modeling.

Table 14. Comparison of modeled univariate continuous odds ratios (OR) for PCA3 and tPSA in matched studies

Study Author, Year[a]	N	PCA3 Regression Coefficients	PCA3 OR @25th Centile	PCA3 OR @75th Centile	Ratio A	tPSA Regression Coefficients	tPSA OR @25th Centile	tPSA OR @75th Centile	Ratio B	Ratio (A/B)
Perdona[31], 2011[c]	218	0.02956	1.77	8.75	4.93	0.22952	3.48	9.0	2.50	1.97
de la Taille[95], 2011[c]	516	0.01980	1.43	3.09	2.16	0.11333	1.74[b]	8.0[b]	1.57	1.38
Aubin[91], 2010 [c]	1,072	0.01882	1.31	2.14	1.63	0.01882	-	-	-	-
Cao[94], 2011 [c]	131	0.07232	-	-	-	0.07232	-	-	-	-

N = number; OR = odds ratio; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
[a]Both studies used in the analysis only included patients in the tPSA "grey zone." Rows 1-3 are shaded.
[b]Ratios A and B are ratios of ORs at 25th and 75th centiles.
[c]These studies reported insufficient information to compute and/or compare Ratios A and B.

Table 15. Measures of independence of PCA3 and tPSA in identifying men with a positive biopsy

Study Author[a], Year	Number	Correlation Coefficient for PCA3 & tPSA in Positive Bx	Correlation Coefficient for PCA3 & tPSA in Negative Bx	PCA3 OR (raw, adj)	tPSA OR (raw, adj)	Other Relevant findings
Ochiai[103], 2011	105	-	-	p<0.001/<0.001	p<0.001/0.52	-
Cao[94], 2011	131	0.079	0.372	1.075/1.055	-	-
Perdona[31], 2011	218	"low"	"low"	1.030/1.030	1.258/1.239	Accuracy improves 3%
de la Taille[95], 2011	516	0.042[b]	0.042[b]	1.020/1.010	1.120/1.150	Accuracy improves 5.5%
Aubin[91], 2011	1,072	-	-	1.019/1.015	1.106/1.087	AUC improves; 69 - 75%

adj = after adjustment for other markers; Bx = prostate biopsy; OR = odds ratio, raw = as observed; PCA3 = prostate cancer antigen 3 gene; tPSA = total prostate specific antigen
[a]Studies ordered by increasing size; rows 3-5 are shaded, indicating studies focusing on the "grey zone" of tPSA.
[b]Correlation coefficient of 0.042 for all biopsy results.

PCA3 and tPSA Elevations: Diagnostic Accuracy

PCA3 and tPSA GRADE Strength of Evidence: LOW

The rationale for "low" follows the GRADE assumption that the high risk of bias in observational studies correlates with a starting strength of evidence of low. The results were deemed to be Consistent, but Indirect. Precision was supported by the ability to observe the expected selection bias of "grey zone" studies and the differences in PCA3 and tPSA performance, but could not be directly measured (e.g., confidence intervals). Strength of association was weak. The results for these domains do not warrant either downgrading to Insufficient or upgrading to moderate.

- Risk of Bias: HIGH
  The quality of individual studies was poor. Three biases were identified that could potentially impact this analysis: partial verification bias, spectrum bias and a sampling bias. Partial verification bias was clearly present for tPSA elevation, but our analyses and a review of the literature indicated that in this setting, the ROC curve was unlikely to be biased (Appendix J). Thus, the focus was towards the ROC and related measurements. Monte Carlo modeling was used to account for the verification bias related to the specific cutoff level at which a certain performance was obtained (Appendix J). Sampling bias was accounted for by stratifying the analyses, when possible. Although there was a relatively high potential for bias to affect select measurements and their interpretation, the measures taken as part of the analyses result in a low risk of those biases influencing the final interpretation. Spectrum bias needs to be considered in addition to the performance estimates (sensitivity, specificity). For example, even though PCA3 has a higher sensitivity at any given specificity, the included studies provided no evidence that those identified as positive with either test had similar or different distribution of disease severity. Although PCA3 appears to be statistically better, it does not necessarily follow that it is clinically superior. Publication bias was informally evaluated and not considered to be an important source of potential bias. However, given the poor quality of all the individual included studies, there is potential for unidentified biases to have occurred.
- Consistency: CONSISTENT
  Overall, analysis showed that PCA3 measurements had higher sensitivity at any specificity compared with tPSA, and higher specificity at any sensitivity. However, it was not possible to formally test for heterogeneity, as original data were not available. No study reported a matched analysis.
- Directness: INDIRECT
  The intermediate outcome of diagnostic accuracy (PCA3 and tPSA) shows both types of indirectness: (1) one body of evidence links the test to the intermediate outcome of diagnostic accuracy and another body of evidence is needed to link the test-related intervention(s) to health outcomes; and (2) based on the lack of matched analyses, it is not possible to determine the extent to which PCA3 and tPSA (or other comparators) are identifying cancer with the same or different characteristics (e.g., aggressiveness) within the spectrum of the disease, and yet another body of evidence is needed to resolve this question.
- Precision: PRECISE

A formal analysis of precision (e.g., confidence intervals) was not able to be computed due to the matched nature of our analyses and the lack of original data. In one analysis that included 20 studies (AUC difference), it was possible to see the reduction in performance for tPSA in a subset of four "grey zone" studies where the AUC difference expanded to a median of 16 percent, compared with the 8.7 percent found in the 16 studies with no sampling bias.

- Strength of Association: WEAK
  Although there is evidence that PCA3 will be slightly better at identifying high risk individuals with a prostate cancer, both PCA3 and tPSA are relatively weak predictors with low sensitivity and low specificity.

PCA3 and tPSA Elevations—Other Intermediate and Long-Term Outcomes

No studies were identified that reported PCA3 and tPSA levels along with specific information on intermediate (impact on decisionmaking about initial or repeat biopsy, biopsy-related harms) or long-term (morbidity/mortality, quality of life, harms) outcomes.

Strength of Evidence: Insufficient

Summary of the Remaining KQs 1 and 2 "Combined" Analyses

Table 16 provides a summary of the numbers of available studies available for each comparator and outcome, as well as the domains (see footnotes) and strength of evidence for each. More detailed descriptions of the data and limited findings for all outcomes and all PCA3 comparators can be found in Appendix K.

KQ 3: Testing for PCA3 and Comparators To Identify Patients with Insignificant Cancer Who May Be Candidates for Active Surveillance

KQ 3 presented a complex clinical scenario. Based on the implementation of tPSA screening and followup, many more prostate cancers are being diagnosed early in the natural history of the disease. The result is the diagnosis of a proportion of cancers that would otherwise not have been diagnosed clinically during the men's lifetimes. Effective risk stratification could inform decisions about whether/when treatment is warranted for such cancers. Alternatively, if risk stratification provides sufficient certainty that the tumor poses little risk to life and health, the patient might benefit from active surveillance and delayed treatment if the disease progresses. The importance of effective schemes for risk stratification was reemphasized by the recent Prostate Cancer Intervention versus Observation Trial (PIVOT)[123] report on 12 year followup of men with histologically confirmed localized prostate cancer (mean age 67 years, stage T1-T2, any grade, tPSA <50 ng/mL). They found no difference in all-cause or prostate cancer-specific mortality between men assigned to observation (watchful waiting) versus those randomly assigned to radical prostatectomy treatment.

The identified studies for KQ 3 investigated the performance of PCA3 and comparators in placing men with biopsy confirmed prostate cancer into categories of clinical risk or significance. Reviewing these studies was complicated by variability in terminology and definitions. Low risk tumors were variably referred to as "low risk,"[94] "indolent,"[89,95,100,117]

"insignificant,"[100,112] or "low volume/low grade."[115,118] High risk tumors were referred to as "intermediate or high risk,"[29,94] "significant,"[95,100,115,118] "unfavorable,"[89] and "aggressive."[118]

Ploussard provides a conceptual definition of insignificant disease as "...a low-grade, small-volume, and organ-confined PCa that is unlikely to progress to clinical and biologic significance without treatment," and that is diagnosed in clinical practice "...in the absence of cancer-related symptoms that would not have caused disease-specific mortality during the patient's life if untreated." [67] Indolent cancers have been characterized as those identified early in the natural history of prostate cancer, possibly prospectively detected by pathologic criteria using tools such as nomograms, and having a good chance of positive outcome with active/aggressive treatment.[14,67] However, these terms have been used interchangeably. We have chosen to use the term "insignificant" to denote the cancers for which active surveillance may be considered.

The first challenge is identifying individuals with insignificant disease who are eligible for active surveillance.[20] The most commonly used criteria used to define insignificant cancer are the Epstein criteria (and modifications).[66] The key prognostic factors are Gleason score ≤6 without Gleason pattern 4 or 5, organ-confined disease (no extraprostatic extension, seminal vesicle invasion or lymph node involvement) and tumor volume less than 0.5 cubic centimeters (cc) (sometimes less than 0.2 cc).[14,66,116] Other criteria may include clinical stage T1c, PSA density less than 0.15 ng/mL/gram, fewer than three positive cores, and less than 50 percent cancer per core.[124] NCCN and others suggests a similar definition for "very low risk."[16] D'Amico low risk criteria are tPSA ≤10 ng/mL, clinical stage T1-T2a, and Gleason score ≤6.[125] This review addresses the potential performance of PCA3 score as a criterion for insignificant disease, but also as a potential marker for an aggressive form of cancer.

The second challenge is determining how to effectively identify progression of disease, to get to a measurable clinical endpoint.[20] How effective is the risk classification system in identifying men with insignificant cancer (clinical sensitivity and specificity)? What are the harms related to misclassification? Answering these questions requires for each risk classification (e.g., insignificant/very low risk, low risk, intermediate risk, high risk) specified measures of progression over time, with and without treatment, as well as assessment of harms and all-cause and prostate-cancer-specific mortality rates.

A validation study of Epstein criteria for insignificant disease in European men found that classification by biopsy criteria "may underestimate the true nature of prostate cancer." At radical prostatectomy, 24 percent of patients with "insignificant" disease had Gleason sum 7-10 scores and 34 percent had non-organ-confined disease.[126] A recent systematic review reported on the accuracy of the Epstein criteria in predicting insignificant prostate cancer.[127] Five of six studies defined insignificance by biopsy criteria and used concordance with prostatectomy pathology to determine accuracy; one study followed biochemical recurrence-free survival for six years. They found significant heterogeneity among the validation studies that was attributed in part to different criteria, variable application of criteria, and changes in the Gleason scoring system. Lack of clinical followup may also be a factor. They concluded that Epstein criteria have suboptimal accuracy for predicting insignificant prostate cancer and require additional, better quality validation studies[127] So, in addition to finding new and most effective markers, better designed validation studies are also needed.

Table 16. KQs: Summary of available studies and strength of evidence for four outcomes and all PCA3 comparators

Table 16a. Available studies, analyses and strength of evidence for the KQ 1 and KQ 2 intermediate outcome of diagnostic accuracy

Comparators	tPSA[a]	%fPSA[a]	PSAD[a]	EVN[a]	Multivariate Models Including tPSA[a]	cPSA[a]	tPSA DT and tPSA Velocity[a]
GRADE: Risk of Bias	High	High	High	High	High	---	---
GRADE: Consistency	Consistent, with 22 studies	Inconsistent[b], with 7 studies	Unknown[b], with 3 studies	Unknown[b], with 4 studies	Unknown[b], with 3 studies	---	---
GRADE: Directness	Indirect	Indirect	Indirect	Indirect	Indirect	---	---
GRADE: Precision	Precise	Imprecise	Imprecise	Imprecise	Imprecise	---	---
GRADE: Strength of Association	Weak	---	---	---	---	---	---
Strength of Evidence (GRADE)[c]	Low	Insufficient	Insufficient	Insufficient	Insufficient	Insufficient	Insufficient
KQ 1 and KQ 2 Area Under the Curve	n=20[22,24-28,31-40,42-45]	n=5[23,25-27,32]	n=3 [22,28,36]	n=3[24,30,37]	0	0	0
Reported Mean/SD	n=8[23,24,26,32,35-37,43]	n=4[22,26,32,37]	n=2[22,36]	0	0	0	0
Performance at a PCA3 cutoff of 35	n=9[22,25,28,31-34,36,41]	n=1[22]	n=2[22,38]	n=1[37]	0	0	0
ROC Curves- Sensitivity/ Specificity	n=14[22,24,25-28,31-38,40]	n=4[23,25,27,32]	n=3[22,28,36]	n=2[36,37]	0	0	0
Regression Analysis	n=2[22,37]	n=3[22,25,37]	0	n=2[24,30]	n=3[22,25,37]	0	0

%fPSA = percent free prostate specific antigen; cPSA = complexed prostate specific antigen; DT = doubling time; EVN = externally validated nomograms; PSAD = prostate specific antigen density; tPSA = total prostate specific antigen;

[a]Corresponds to KQ 1 and KQ 2.

[b]Consistency could not be assessed due to insufficient data from comparable studies, or because studies did not report results in a consistent manner.

[c]GRADE assessment of strength of evidence for each outcome for each comparator is based on assessment of the evidence for four domains: risk of bias; consistency of effect size/direction, directness of the evidence-health outcome link; and precision (degree of certainty) of effect estimate. Based on the domains, GRADE strength of evidence categories are Insufficient, Low, Moderate and High.

Table 16b. Available studies and strength of evidence for other KQ 1, KQ 2, and KQ 3 intermediate and long-term outcomes

PCA3 Comparators	tPSA Elevations SOE[a] (N[b])	% fPSA SOE[a] (N[b])	PSA Density SOE[a] (N[b])	Externally Validated Nomogram SOE[a] (N[b])	tPSA Velocity/ Doubling Time SOE[a] (N[b])	cPSA SOE[a] (N[b])
KQ 1 & 2 intermediate outcome:Impact on biopsy decisionmaking	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)
KQ 1 & 2 intermediate outcome: Biopsy-related harms	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)
KQ 1 & 2 long-term outcome: Morbidity, mortality, function, quality of life	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)
KQ 1 & 2 long-term outcome: Treatment-related harms	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)
KQ 3 intermediate outcome: Diagnostic accuracy[c]	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)
KQ 3 intermediate outcome: Impact on decisionmaking	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)
KQ 3 intermediate and long-term outcomes: Treatment-related harms	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)
KQ 3 intermediate and long-term outcomes: Health outcomes and surrogates (*e.g.*, biochemical recurrence)	Insufficient[c] (2)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)	Insufficient (0)

%fPSA = percent free prostate specific antigen; cPSA = complexed prostate specific antigen; EVN = externally validated nomograms; PSAD = prostate specific antigen density; SOE = GRADE strength of evidence; tPSA = total prostate specific antigen

[a]GRADE assessment of strength of evidence for each outcome for each comparator is based on assessment of the evidence for four domains: risk of bias; consistency of effect size/direction, directness of the evidence-health outcome link; and precision (degree of certainty) of effect estimate. Based on the domains, GRADE strength of evidence categories are Insufficient, Low, Moderate and High.

[b]N=Number of studies.

[c]Risk of bias: High; Consistency: Unknown; Directness: Direct; Precision: Imprecise.

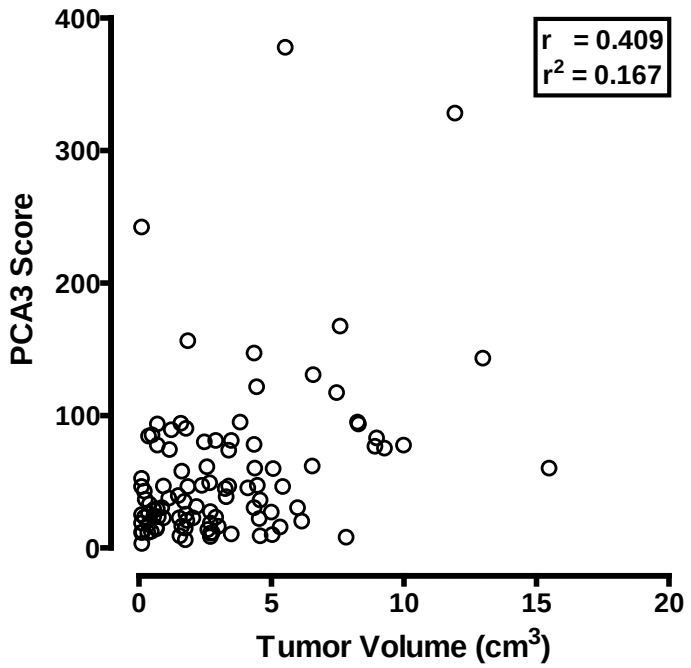
Description of Included Studies

The inclusion criteria for KQ 3 were also set to select only matched studies. These are defined as studies that provide estimates of test performance or other outcomes for PCA3 and at least one other comparator using the same sample set. Studies of PCA3 alone, or of other comparators without PCA3, were, therefore, excluded. Thirteen studies were identified that addressed KQ 3 and reported on PCA3 and other preoperative/pretreatment markers for stratifying tumors by risk (Table 17).[88,89,94,95,100,112-119] Two studies based analyses on biopsy markers without prostatectomy[94,95] and eight reported prostatectomy results as an endpoint.[88,100,112,114-116,118,119] Two studies were conducted on subjects with longitudinal data including short-term followup.[88,89] Tables 1 through Table 4 include descriptive information about these studies. Table 17 provides information and results. Table 18 provides detailed information about the wide variety of markers investigated in these studies for association with low and high risk disease.

Prostatectomy is not useful as an endpoint for determining diagnostic accuracy because it is not a clinical outcome, but rather an intermediate step. Pathological testing of prostatectomy specimens adds data to further assess the tumor as high or low risk, including tumor volume, prostatectomy Gleason score and stage, possible upgrading from biopsy, and other pathological findings (e.g., extracapsular extension, perineural invasion, positive surgical margins). However, the association of PCA3 with these markers, or the ability of PCA3 to predict them at prostatectomy, relates to the determination of risk category, but does not provide the formal evidentiary link between the risk assigned and specific intermediate and long-term clinical outcomes. Without even short-term specified clinical endpoints or validated surrogates, these data cannot be used to provide estimates of diagnostic accuracy.

The included articles address a wide range of comparators (Table 18), many different combinations of criteria defining individuals with "low" or "high" risk prostate cancers, and varying presentations and analyses of the data. However, one result was most consistently reported, and that was an association between PCA3 score and tumor volume (Table 17). Most studied PCA3 and comparators as potential predictors of insignificant cancer, while others reported possible use for identifying aggressive cancer.[116,117] Three studies reported that PCA3 was an independent predictor of tumor volume, though cutoffs and endpoints (greater than or less than the 0.5cm^3 tumor volume cutoff) differed.[113,116,119] Five studies[113,115,116,118,119] reported higher correlations between PCA3 and tumor volume ($r = 0.27$ to 0.41; $p \leq 0.04$) than for other comparators (e.g., tPSA, %fPSA, PSA density, clinical stage, biopsy Gleason score). Unfortunately, those correlations may be suspect. Most studies did not provide scatterplots of the data. The two that did[89,116] clearly show that the data for both PCA3 and tumor volume should be subject to transformation prior to computing the correlation. Figure 12a redisplays data published by Ploussard.[116] These data were digitized from a provided figure and should be considered reasonably accurate, but not as reliable as original raw data. This analysis is aimed at demonstrating a more appropriate analytic methodology. The correlation is lower after transformation (but still significant). However, examining the data (Figure 12b), it appears that most of the "prediction" is confined to very high PCA3 scores (>120) associated with very large tumors (>2 cm^3).

Figure 12a. Scatterplot illustrating the correlation between PCA3 scores and tumor volume in prostatectomy specimens on a linear scale (data from Ploussard et al., 2011[116])



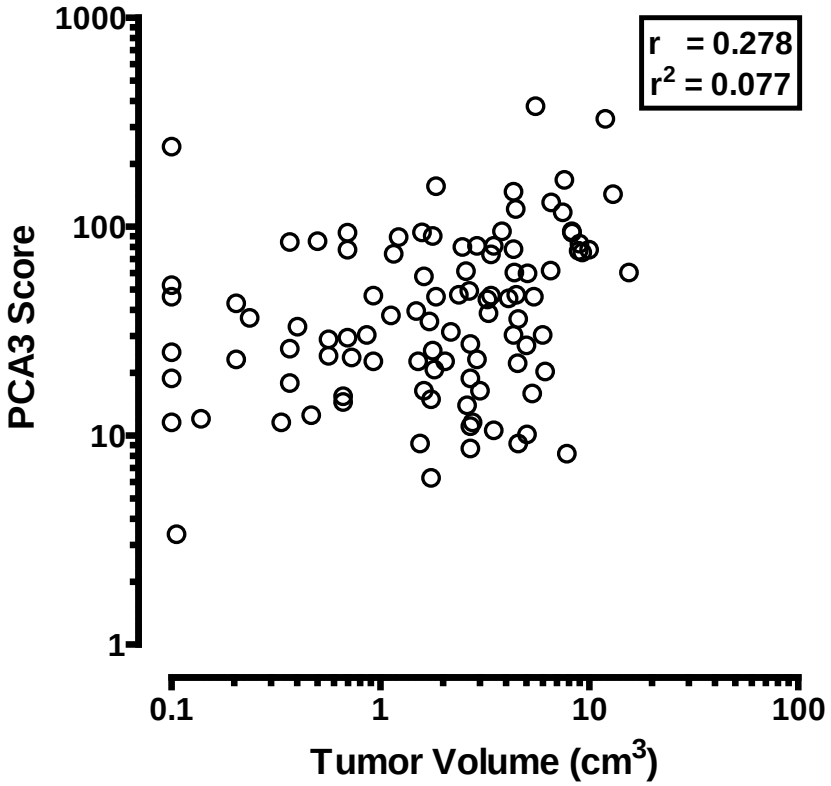
PCA3 = prostate cancer antigen 3 gene
Note: For each subject, PCA3 score is shown on the x-axis and tumor volume in cubic centimeters (cm3) on the y-axis. The highly significant r-squared value ($p<0.001$) indicates that about 17% of the variability in tumor volume can be accounted for by the range of PCA3 values. However, both measurements (tumor volume and PCA3 score) are highly right-skewed, violating the assumption of being a Gaussian distribution required for the computation of reliable correlation coefficients.

Only two of the studies reviewed for this report had a longitudinal component and described a clinical outcome other than pathological results of prostatectomy.[88,89]

- Lymph node involvement in a prostate cancer patient is an indicator of poor clinical outcome. One study[88] attempted to identify "micrometastases," based on identifying tumor cells within the lymph nodes that are producing the prostate cancer markers tPSA and PCA3. The method used to quantify these markers in lymph node extracts was real time reverse transcriptase PCR (RT-PCR) for both PCA3 and PSA mRNA. The study followed 120 patients with localized prostate cancer for 4 to 6 years and used biochemical recurrence (any serum tPSA greater than 0.2 ng/mL) as the surrogate outcome of interest. As expected, they found significantly decreased biochemical recurrence free survival among the 11 subjects with histologically confirmed lymph node metastases, compared with 77 subjects with no lymph node involvement. Among the remaining 32 patients with biochemical recurrence, many were identified as having micrometastases based on either tPSA or PCA3 (or both) testing. tPSA had a sensitivity for biochemical recurrence of 73 percent and a false positive rate of 22 percent ($p<0.001$). PCA3 had a lower detection (42 percent) and a comparable false positive rate (23 percent), but the effect was not significant ($p=0.095$). While this appears to indicate that PSA testing is more predictive, the use of PSA mRNA as the test, and a rise in serum tPSA levels as the outcome suggests an important risk of bias. The authors provided no

information on validation of quantitative testing for these biomarkers in this sample type or on confirmation of the results using a published method.

Figure 12b. Scatterplot illustrating the correlation between PCA3 scores and tumor volumes in prostatectomy specimens after logarithmic transformation (reanalysis of data from Ploussard et al., 2011[116])



PCA3 = prostate cancer antigen 3 gene
Note: For each subject, PCA3 score is shown on the logarithmic y-axis and log tumor volume in cubic centimeters (cm3) on the logarithmic x-axis. Both distributions are now reasonably Gaussian and the corresponding r-squared value is reduced to just under 8% of the total variability. This is about half of the value found prior to transformation. In addition, the corresponding p-value is reduced to 0.004; still highly significant.

- Based on no more than 2-year followup of patients in an active surveillance program, Tosoian et al. reported PCA3 and tPSA results (mean, standard deviation, median) for the 38 of 294 patients progressing to treatment based on yearly biopsy results.[88,89] Epstein criteria were used for initial enrollment in the surveillance program. Progression to treatment was recommended for "unfavorable" findings, defined as any Gleason pattern 4 or 5, greater than 2 positive biopsy cores, or more than 50 percent involvement of any core with cancer. No difference in PCA3 and tPSA levels was observed between the 13 percent who progressed and those remaining in active surveillance (p=0.13). However, the authors state that only 140 of the 294 study subjects submitted a urine sample, and did not report how many of these 140 men had an unfavorable result on biopsy. This study did not provide matched results for all subjects (partially matched).

No studies were identified that reported on other intermediate outcomes (e.g., diagnostic accuracy, decisionmaking, harms) or long-term clinical outcomes (e.g., mortality/survival, morbidity, quality of life). All studies were judged to be poor quality, mainly due to lack of clinical followup, but also to lack of information on study subjects. Six studies were funded by GenProbe and six disclosed authors with potential conflicts of interest (Table 1); others did not report on source of funding or conflicts of interest (Table 1). The detailed results of assessment of quality of individual studies addressing KQ 3 are presented in Table F-2 in Appendix F.

PCA3 and Comparators—Intermediate Outcome: Diagnostic Accuracy

- Risk of Bias: HIGH
  The quality of individual studies was poor. All studies were observational, raising a high potential for biases to have occurred.
- Consistency: UNKNOWN
  No studies were identified that reported on matched data for PCA3 and comparator results, and also reported specific clinical outcomes of patients with tumors characterized as low risk and high risk, who:
  - opted for active surveillance and never progressed to treatment;
  - opted for active surveillance and progressed to treatment; or
  - opted for immediate treatment.

  No effect(s) could be measured.
- Directness: DIRECT
  This should be Direct, as the evidence would ideally determine diagnostic accuracy by linking the risk assignment based on testing/pathological results directly to health outcomes.
- Precision: IMPRECISE
  This cannot be assessed, as no comparisons were possible based on the two studies of different populations, using different assays, and reporting different surrogate outcomes.

Strength of Evidence: Insufficient

Strength of evidence could not be evaluated. Only two studies were identified, and they did not perform the studies in the same setting, have the same sample type or have comparable outcome measures.

PCA3 and Comparators—Intermediate Outcome: Impact on Decisionmaking

No studies were identified that reported PCA3 and comparator results and intermediate outcome data (e.g., physician or patient surveys, chart review) on the degree to which PCA3 or comparator test results and categorization of risk as high or low impacted decisions made with regard to selection of active surveillance versus aggressive treatment.

Strength of Evidence: Insufficient

PCA3 and Comparators—Intermediate and Long-Term Outcome: Treatment-Related Harms

Studies have been conducted that document treatment-related clinical harms such as incontinence and impotence. Based on general studies on potential psychosocial harms of diagnostic testing, it is possible to generalize that patients facing treatments such as radical prostatectomy might also experience anxiety or perceive a reduction in quality of life. However, no studies were identified that reported PCA3 and comparator test results and intermediate outcome data (e.g., physician or patient-reported adverse events, biochemical recurrence, progression to treatment) on the degree to which categorization of risk as high or low and choice of active surveillance or treatment related to the occurrence of adverse clinical events.

Strength of Evidence: Insufficient

PCA3 and Comparators—Intermediate and Long-Term Health Outcomes

No studies were identified that reported PCA3 and comparator results and the association of low and high risk categorization with long-term outcomes such as mortality/survival and morbidity (e.g., function, quality of life) of the selected course of management or treatment. However, two poor quality studies reported on relatively short-term health outcomes, biochemical recurrence and progression from surveillance to treatment in an active surveillance program.

Strength of Evidence: Insufficient

Table 17. Characteristics of matched studies addressing KQ3 with biopsy and prostatectomy results

Author, Year	Biopsy GS, %	Clinical Stage,%	Number of P	P Gleason Scores, %	Pathologic Stage, %	"Low Risk" Prostate Cancer Category, %	Conclusions – PCA3:
Auprich, 2011[112]	<7, 50.8 ≥7, 49.2	-	305	< 7, 27.9 ≥ 7, 72.1	NR	Insignificant[a], 10	Median scores lower with low TV and insignificant PrC[b] (p<0.001); improves multivariate AUC
Cao, 2011[94]	≤ 6, 50 = 7, 38 ≥ 8, 12	≤T2a, 25; T2b, 52; T3, 23	-	-	-	Low risk[b], 8.1	Has significant correlation with low risk group; no RP
de la Taille, 2011[95]	< 6, 1 = 6, 52 = 7, 42 > 7, 5	T1c, 86; T2, 13; T3, 1	-	-	-	Indolent[c], 25	Median scores higher with GS ≥7, % positive cores > 33%, and significant PrC; no RP
Durand, 2012[113]	= 6, 43.1 = 7, 50.6 > 7, 6.3	T1a/b, 0.6 T1c, 81.8 T2, 18.2	160	= 6, 43.1 = 7, 50.6 ≥ 8, 6.3	≤ pT2, 70.6 pT3, 27.5 pT4, 1.9	NR	Score >35 correlates with (r=0.34, p < 0.01) and predicts (OR 2.7;p=0.04) TV; predicts positive surgical margins (OR 2.4, p=0.04) and GS >6 (p<0.001)
Hessels, 2010[100]	≤ 6, 74.3 = 7, 22.9 ≥ 7, 2.8	NR	70	<7, 43 ≥7, 57	pT2, 59.0 p T3, 42.0	Indolent[d] / Insignificant, 8.5	Predictive value for aggressiveness features not confirmed in this study
Kusuda, 2011[88]	-	-	120	≤ 6, 56.7 = 7, 33.3 ≥ 8, 10.0	pT2, 55.8 pT3, 42.5 pT4, 1.7	NR	PCA3 expression in LN tissue can predict biochemical recurrence-free survival after prostatectomy
Liss, 2011[114]	≤ 6, 53.1 > 6, 47.9	NR	98	≤ 6, 29.9 > 6, 71.1	pT2, 77.3 pT3a, 16.5 ≥pT3b, 6.2	NR	Association with PNI (p=0.05), not pathological stage, GS > 6, or EPE
Nakanishi, 2008[115]	= 6, 40.6 = 7, 53.1 = 8, 6.3	T1c, 70.8 ≥ T2, 29.2	96	= 6, 15.6 = 7, 77.1 = 8, 3.1 = 9, 4.2	pT2, 82.3 pT3a, 10.4 pT3b, 7.3	Low volume / low grade[e], 11	Median scores higher for low volume/low grade vs. significant PCa (p=0.007); correlation with TV (p=0.01)
Ploussard, 2011[116]	= 6, 100		106	= 6, 58.4 = 7, 41.6	≤ pT2, 76.4 pT3, 23.6	Low risk cohort[f] in active surveillance	Correlates with (r=0.41, p<0.001) and at > 25 predicts (p=0.01) TV and insignificant PrC (p=0.02)

Table 17. Characteristics of matched studies addressing KQ3 with biopsy and prostatectomy results (continued)

Author, Year	Biopsy GS, %	Clinical Stage,%	Number of P	P Gleason Scores, %	Pathologic Stage, %	"Low Risk" Prostate Cancer Category, %	Conclusions – PCA3:
Tosoian, 2010[89]	< 7, 57.9 ≥ 7, 42.0	NR	-	-	-	Low risk[g] PrC patients in active surveillance	Score was not significantly associated with short-term biopsy progression.
van Poppel, 2011[117]	≤ 6, 53 = 7, 43 > 7, 4	T1c, 79 T2-T2c, 21 T3a, 1	175	< 7, 32.1 ≥ 7, 67.9	T2a-c, 78.6 T3a-b, 21.4	Indolent[h], 21.7	Median scores lower with GS <7 (p<0.001) and stage pT2a-pT2c (p=0.01)
Vlaeminck-Guillem, 2011[118]	= 6, 47 = 7, 45 = 8/9, 8	T1, 84 T2, 16	102	NR	NR	Low volume / low grade[j], 8.2	Score correlated with TV (p< 0.001), multifocality (p = 0.012), and apical / basal invasion (p<0.05)
Whitman, 2008[119]	= 6, 69.4 = 7, 20.8 = 8/9, 9.7	T1, 71.2 T2, 28.8	72	= 6, 58.3 = 7, 31.9 = 8/9, 9.7	pT2, 70.9 pT3a, 20.8 pT3b, 8.3	-	Correlated with TV (r=0.38, p < 0.01) and was an independent predictor of TV <0.05 (p=0.04) and ECE (p=0.01); improves multivariate AUC

AUC = area under the curve; ECE = extracapsular extension; GS = Gleason score; LN = lymph node; NR = not reported; P = prostatectomy; PCA3 = prostate cancer antigen 3 gene; PNI = perineural invasion;
SVI = seminal vesicle invasion; TV = tumor volume

[a]Defined by Epstein criteria: organ confined, tumor volume less than 0.5 cc, absence of Gleason grade 4 or 5.

[b]Defined as tPSA ≤ 10 ng/mL, GS ≤ 6, clinical stage ≤ T2a.

[c]Defined by Epstein criteria: stage T1c, PSAD < 0.15 ng/mL, biopsy GS ≤ 6, percent positive cores ≤ 33%.

[d]Defined as organ confined, TV < 0.5 mL and absence of Gleason grad 4 or 5 disease.

[e]Defined as organ confined, dominant tumor volume less than 0.5 cc, and absence of Gleason grade 4 or 5.

[f]Defined as tPSA ≤ 10, clinical stage T1c-T2a, GS 6 prior to radical prostatectomy.

[f]Defined by Epstein criteria: stage T1c, PSAD < 0.15, biopsy GS ≤ 6, ≤ 2 positive cores, and no more than 50% involvement of any one core.

[h]Defined by Epstein criteria: stage T1c, PSAD < 0.15 ng/mL, biopsy GS ≤ 6, percent positive cores ≤ 33%.

Organ confined cancer (pT2 or less) with a total tumor volume of less than 0.5 ml, and absence of Gleason grade 4 or 5 disease (contemporary Epstein criteria).

Table 18. Comparators investigated with PCA3 scores in matched studies addressing KQ 3

Numbers of Subjects and Reported Comparators	Auprich 2011[112]	Cao 2010[94]	de la Taille 2011[95]	Durand 2012[113]	Hessels 2010[100]	Kusuda 2011[88]	Liss 2011[114]	Nakanishi 2008[115]	Ploussard 2011[116]	Tosoian 2010[89]	van Poppel 2011[117]	Vlaeminck-Guillem 2011[118]	Whitman 2008[119]
No. of biopsy patients	-	131	515	-	336	-	-	-	-	294	348	-	-
No. of prostatectomy patients	305	-	-	160	70	120	98	96	106	-	175	102	72
Biomarkers	-	-	%fPSA, PSAD	-	-	-	PSAD	-	%fPSA, PSAD	-	%fPSA, PSAD	%fPSA	-
Biopsy: Gleason score	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Biopsy: Clinical stage	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes
Biopsy: Localized/organ confined PCa	Yes	NR	NR	Yes	NR	Yes	Yes	NR	Yes[d]	Yes[f]	NR	NR	NR
Biopsy: Percent tumor positive cores	Yes	No	Yes	Yes	No	No	No	Yes	Yes	No	Yes	No	No
Biopsy: Percent tumor per core	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Biopsy: Prostate volume	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No
Prostatectomy: Gleason score	Yes	NA	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Prostatectomy: Pathological stage	No	NA	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Prostatectomy: Tumor volume	Yes	NA	No	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes
Prostatectomy: Multifocality	No	NA	NA	No	No	No	No	No	No	No	No	Yes	No
Prostatectomy: Extracapsular extension	Yes	NA	NA	No	No	No	Yes	No	No	No	No	No	Yes
Prostatectomy: Seminal vesicle invasion	Yes	NA	NA	No	No	No	No	No	No	No	No	No	No
Prostatectomy: Perineural invasion	No	NA	NA	No	No	No	Yes	No	No	No	No	Yes	No

Table 18. Comparators investigated with PCA3 scores in matched studies addressing KQ 3

Numbers of Subjects and Reported Comparators	Auprich 2011[112]	Cao 2010[94]	de la Taille 2011[95]	Durand 2012[113]	Hessels 2010[100]	Kusuda 2011[88]	Liss 2011[114]	Nakanishi 2008[115]	Ploussard 2011[116]	Tosoian 2010[89]	van Poppel 2011[117]	Vlaeminck-Guillem 2011[118]	Whitman 2008[119]
Prostatectomy: Surgical margin status	No	NA	NA	Yes	No	Yes	No	No	No	No	No	Yes	No
Risk criteria: Epstein	Mod[a]	No	No	No	No	No	No	No	Yes[e]	Yes[f]	Yes[g]	No	No
Risk criteria: Other (see footnote)	No	Yes[b]	No	No	Yes[c]	No	No	No	No	No	No	Yes[h]	No
Lymph node mRNA expression	NA	NA	NA	NA	NA	Yes[i]	NA	NA	NA	NA	NA	NA	NA
PCa progression	No	No	No	No	No	Yes[i]	No	No	No	Yes[j]	No	No	

%fPSA = percent free prostate specific antigen; Mod = modified Epstein criteria; NA = not applicable; NR = not reported; PCa = prostate cancer; PSAD = total prostate specific antigen density; RP = radical prostatectomy; TV = tumor volume

[a] *Insignificant PCa* = organ-confined PCa, TV < 0.5 mL, GS < 7.

[b] *Low risk PCa* = tPSA ≤ 10 ng/mL, GS ≤ 6, clinical stage ≤ T2a; *high risk PCa* = tPSA >20, GS ≥ 8, clinical stage ≥ T3a.

[c] *Insignificant PCa* = low tumor volume/low grade.

[d] Study enrolled only patients with "low risk" PCa, defined as tPSA ≤ 10 ng/mL, clinical stage T1c-T2a and biopsy GS = 6.

[e] *Insignificant PCa* = organ-confined, no Gleason pattern 4 or 5, TV < 0.5 mL.

[f] Study enrollment criteria were based on the Epstein criteria: clinical stage T1c, PSA density < 0.15 ng/mL/cm^3, GS ≤ 6, 2 or less biopsy cores with cancer and maximum 50% involvement in any one core.

[g] *Indolent PCa* = clinical stage T1c, PSA density < 0.15, biopsy GS ≤ 6, % positive cores ≤ 33%.

[h] *Significant PCa* = RP GS ≥ 7, pathological stage pT3 or greater, TV ≥ 0.5 mL.

[i] Biochemical recurrence based on lymph node metastasis identified by PCA3 RNA expression in lymph node tissue.

[j] Progression to treatment in an active surveillance program, defined by surveillance biopsy with any Gleason pattern 4 or 5, >2 positive cores, or more than 50% involvement of any core with cancer.

Discussion

KQ 1 and KQ 2: Findings and Strength of Evidence

This Comparative Effectiveness Review investigated three Key Questions (KQs). The first two KQs addressed the performance of PCA3 testing in comparison with six individual serum biomarkers, or combinations of such biomarkers, to predict risk of prostate cancer at biopsy among men identified as being at risk through tPSA screening. These men were candidates for initial biopsy (KQ 1) or repeat biopsy after one or more previous negative biopsies (KQ 2). Findings were limited by several factors, including the relatively small number of matched studies, that no studies reported raw data or a true matched analysis, that clinical follow-up was not performed in essentially all studies identified for either of the KQs, and that the individual studies were of poor quality.

The only comparator with sufficient data for analysis was tPSA. However, this analysis was subject to three important biases: (1) sampling bias, due to some studies limiting enrollment to men in the "grey zone" of tPSA elevations (e.g., 2.5 ng/mL to less than 10 ng/mL); (2) verification bias, due to the use of tPSA elevations in the decision-making process to accept (or reject) prostate biopsy; and spectrum bias; and (3) spectrum bias. Sampling bias was addressed by separately reporting, when possible, studies in the "grey zone" from those without tPSA level restriction. Verification bias was addressed by reviewing relevant publications and performing in-house modeling. The spectrum effect of most concern could only be described. It relates to the range of severity of disease identified by PCA3 and comparators. In order to estimate true clinical validity, it is necessary to examine both concordant and discordant test results (those positive by one test but negative by the other) in men with positive biopsies. If the men with discordant results showed similar severity of disease, the sensitivity/specificity estimates would be both statistically and clinically equivalent. However, if one test identifies more severe cases that the other misses, the estimates may be statistically equivalent but clinically different.

The main findings included: PCA3 consistently has higher sensitivity than tPSA elevations at a given specificity, PCA3 and tPSA elevations provide essentially independent information about prostate cancer risk, and the performance of PCA3 relative to tPSA is not dependent on the biopsy status (i.e., the relative performances are not significantly different between men having initial and men having repeat biopsy). However, the overall strength of evidence for these findings is Low. This information could be of potential use to researchers designing future studies and may also be of interest to policy-makers with regards to future testing strategies for better stratification of prostate cancer risk in men.

Potential To Combine Studies Addressing KQs 1 and 2

Given the limited data available for KQ 1 and KQ 2, we examined the pool of studies that provided comparison data, but were not limited to only men having an initial biopsy or men with no previous positive biopsy having a repeat biopsy. An additional 16 studies provided usable comparative data for PCA3 and tPSA, but the enrolled population consisted of a mixture of men in both categories or did not report this proportion. The proportion of enrolled men having an initial biopsy (100 percent would indicate a study suitable for KQ 1, while 0 percent would indicate a study suitable for KQ 2) was then plotted against two separate measures of relative effect size (difference in AUC and sensitivity at selected specificities). In both analyses, the correlation was essentially zero, indicating that neither of the two relative effect size measures

was dependent on the mixture of initial versus repeat biopsies. This provided sufficient evidence to perform a series of "combined" KQ 1/KQ 2 analyses.

KQ 3: Findings and Strength of Evidence

The third question addressed the performance of PCA3 testing in comparison with serum biomarkers and, other risk factors (e.g., family history, age) and pathological tumor markers (e.g., Gleason score, staging) to identify men with high risk (i.e., aggressive) and low risk (i.e., insignificant/indolent) prostate cancers. KQ 3 focused on matched studies of men with a positive prostate biopsy. In order to inform decisions about management and treatment options (i.e., active surveillance vs. treatment). KQ 3 investigated risk assessment and the potential to categorize patients based on specific biomarker results and pathological markers from biopsy and, mainly, prostatectomy.

However, the reference standard for diagnostic accuracy must be a longer term clinical endpoint, in order to investigate outcomes in the context of categorization of risk. These endpoints might include measures of progression, metastasis, and prostate cancer related morbidity (e.g., function, quality of life) or mortality. Only two studies provided information on progression related to PCA3 and comparators, but in different populations with poorly described surrogate clinical endpoints. More time will be needed for assessment of progression free survival, mortality and other long-term outcomes.[20] Given the relatively recent advent of PCA3 testing, it is not surprising that no studies were identified that provided intermediate or long term outcomes based on PCA3 and one or more comparators.

Findings in Relationship to What Is Already Known

Systematic Review of PCA3

One prior structured systematic literature review of PCA3 testing and prostate cancer was identified.[80] That review covered the time period 2000 to 2009 and included only studies of diagnostic accuracy, with prostate biopsy as the gold standard. No comparisons were made to other biomarkers. Included were studies reporting on cohorts of adult men undergoing prostate cancer screening; mean tPSA levels were provided for each study, but elevated tPSA was not an inclusion criterion. Fourteen studies were included in the Ruiz-Aragon review[46,47,82-86,96,104,115,128-130] The four studies from our literature search that were also reviewed by Ruiz-Aragon were excluded from our review due to high likelihood of data duplication with other included studies.[47,82-84]

Summary estimates of clinical sensitivity and specificity from the previously published meta-analysis (random effects model) were 63 percent (95% CI: 60-66 percent) and 75 percent (95% CI: 73 to 76 percent), respectively. Heterogeneity was high and significant ($p < 0.001$). The SROC AUC was 0.783. The authors acknowledged that no consensus existed on the most appropriate PCA3 cutoff for clinical decisionmaking. Based on QUADAS criteria, the authors reported the studies to be of moderate to high quality, in spite of the acknowledged lack of blinding. However, the review did not compare PCA3 performance with any other biomarkers, used reported sensitivity/specificity results from studies with varying PCA3 cutoffs (range, scores of 19 to 66), included studies using probability cutoffs, and included a study of prostatic fluid samples. These inclusion criteria may account for a portion of the observed heterogeneity.

Applicability

KQs 1 and 2: Biopsy Decisionmaking

To determine the effectiveness of PCA3 and comparator tests in predicting risk of prostate cancer at initial or repeat biopsy, or in risk categorization to inform decisions about treatment, it is useful to assess the applicability of the findings in this review.

Population and Settings of Care

The populations studied in these studies were largely drawn from convenience samples at academic medical centers where patients with elevated tPSA results and/or other risk factors (e.g., positive DRE, family history, African American) seek referral or specialty care. Such observational cohort studies are subject to spectrum and verification biases. For example, these studies may represent a group of men at higher risk of prostate cancer, or risk of men with a difference in severity of disease, than the total cohort of men with elevated tPSA and/or other risk factors. If a study reports results in which biopsy is tPSA-related, verification bias can impact the accuracy of the sensitivity and specificity estimates at select tPSA cutoffs. If those not accepting biopsy are considered missing, this is considered "partial verification" bias. However, no studies addressed these potential biases.

The positive biopsy rate in such referral populations will depend on multiple factors, including the tPSA cutoff, the number of men with elevated tPSA who opt out of biopsy (e.g., men with lower tPSA levels and lower risk), and/or the proportions of men with other important risk factors. In 17 included studies, biopsy positive rates ranged from 16.9 to 72.9 percent, with a median of 36 percent. It is unlikely that the clinical sensitivity and specificity estimates derived from these studies will be affected. However, positive and negative predictive values do depend on disease prevalence, and the reported predictive values will vary. Total PSA is currently the standard initial screening test for the identification of men at increased of prostate cancer. However, studies of proficiency testing data have shown that tPSA test kits provide variable results that could result in significant differences in proportion of identified cancers at recommended tPSA cutoffs. Therefore, an inherent limitation or source of heterogeneity for the PCA3 test may be the use of PSA for first order screening.

The observation that the relative performance of PCA3 versus tPSA elevations does not appear to be dependent on the biopsy history is a new finding. The current FDA approval for PCA3 restricts it use to decision-making regarding a repeat biopsy in men with a specific clinical history that includes one or more negative prostate biopsies. This review suggests that PCA3 might also have use in men making decisions about an initial biopsy. This would greatly broaden the settings in which PCA3 might be offered. However, the data supporting the FDA approval have not been published and were not available for this evidence review. The data provided in the FDA summary were specific to repeats and could not be included in key analyses because it was not a matched study design.

Whether or not further utilization of PCA3 occurs will also involve other important dynamics, such as acceptability and costs. Further studies will also be needed to validate models incorporating PCA3 testing with other markers in the setting of the initially positive patient. Uptake of PCA3 testing by physicians could be affected by other factors. The FDA approval of the PCA3 test and the Prostate Health Index test may be seen as sources of additional information that could improve tPSA performance. The current controversy around the utility of tPSA screening has become more vocal since the recent U.S. Preventive Services Task Force

concluded that evidence of harms to men outweighed the evidence of benefits and recommended against screening men in the general population, regardless of age, until there is a better test. In addition, it is unclear how acceptable the PCA3 test is to men, since it involves prostate massage and urine collection, as opposed to a simple blood draw.

Interventions

No publications were identified that addressed the impact of adding PCA3 scores to the process of making decisions about proceeding to initial or repeat biopsies. Therefore, uncertainty remained about how the test would, or should, impact practice (e.g., all screen positive men, men in the "grey zone," only for repeat biopsy decisions). For example, some may argue that clinicians could intervene based solely on a very elevated tPSA, but use additional markers to evaluate the remaining "grey zone" patients. This assumes that very elevated tPSA results are, by themselves, sufficiently informative for decisionmaking, and performance would not benefit from adding a second useful and independent marker like PCA3. Should PCA3 come into routine practice, it is not clear that use only in the "grey zone" would be an effective approach. Other issues that need clarification include: how the testing would be integrated into protocols for management of men with elevated tPSA or other risk factors, and whether physicians and patients would be receptive to using PCA3 results in the biopsy decision.

Comparisons

For this intended use, analysis of matched studies showed that PCA3 had slightly higher performance compared with the extent of tPSA elevations. This result is based on a summary analysis of matched within-study differences. Some studies reported the sensitivity and specificity for PCA3 and tPSA, but no studies actually reported a full matched analysis. A full matched analysis might provide 2x2 tables (or raw data for analysis) that allowed for direct comparison of PCA3 and a comparator in biopsy positive and biopsy negative samples (i.e., identifying those that both tests called positive, those both called negative, and those on which results were inconsistently called).

Studies of PCA3 and comparators other than tPSA (e.g., %fPSA, PSA density, PSA velocity or doubling time, complexed PSA, and externally validated nomograms) were inadequate or completely lacking. Even if PCA3 were a better secondary test than tPSA elevations (i.e., detected the same number of cancers, but with fewer biopsies in men without cancer), no data have shown that the identified men have the same type of prostate cancer (e.g., aggressive, insignificant/indolent). Fully matched analyses of existing raw data could help answer this question.

Although the focus is generally on comparing PCA3 with other prostate cancer markers, neither PCA3 nor tPSA would generally be considered to have high screening performance. To improve overall performance, both biomarkers as well as demographic information have been combined to predict the risk of prostate cancer. A multivariate patient-specific risk could be derived to be used in personalized decisionmaking regarding the benefits and harms of prostate biopsy.

Outcomes

Applicability was limited by the lack of information on the impact that reporting PCA3 scores in a clinical setting might have on intermediate and long-term outcomes. Of primary interest is whether patient decision-making regarding prostate biopsy would be impacted by

reporting PCA3 results (improved diagnostic accuracy), and whether those decisions might result in higher positive predictive values (fewer biopsies performed in men without prostate cancer). Future studies might consider routinely combining the tPSA elevations with PCA3 results according to one of several validated nomograms contained in this review and quantifying these effects.

The population could be randomized into those with PCA3 included versus those receiving equivalent routine care without PCA3 testing being performed for clinical use. Both groups would have urine samples collected; one group would be immediately tested, while the other is tested at a later time. Cutoffs levels could be selected such that the sensitivities are similar in both groups, but fewer biopsies would be expected with the use of PCA3. This would be expected to result in a higher positive predictive value. Once the study is complete, patients could receive additional results as if they were had been randomized to the other group. This would allow for a fully matched study in which cancers were detected by both, as well as comparing the cancer detected by one or the other.

The remaining intermediate and longer term outcomes are more challenging due to the difficulty of collecting this information and the lack of validated surrogate measurements.

KQ 3: Management of Biopsy Positive Patients

To determine the effectiveness of PCA3 and comparator tests in identifying patients who are candidates for active surveillance versus therapy, it is useful to assess the applicability of the findings in this review in a systematic manner, reflecting both on the design and execution of the included studies. However, no studies reported clinical outcomes of patients classified as low risk and enrolled in active surveillance versus those of high risk patients for whom treatment was recommended or elected. Two poor quality longitudinal studies reported on PCA3 and tPSA, but data could not be compared or combined. Detection of micrometastases in lymph node tissue as a predictor of biochemical recurrence-free survival was outside the inclusion criteria because no tPSA serum testing or PCA3 urine testing was performed along with lymph node extracts. The study of two-year followup on progression of patients from active surveillance to treatment had design flaws (e.g., partially matched, incomplete data).

Implications for Clinical and Policy Decisionmaking

The published literature on the use of PCA3 and comparators in the two intended uses described in KQ 1 and KQ 2 was found to be limited and of poor quality. However, the recent FDA approval of the Gen-Probe PCA3 test for the intended use addressed in KQ 2 will raise awareness of this test, and possibly accelerate its adoption into practice. Practice guidelines currently recommend that a decision to have tPSA testing should be based on discussion between the physician and patient on the balance of potential benefits and harms. An increase in the knowledge base on the comparative effectiveness of PCA3 and other biomarkers is needed to support more informed choices.

The pros and cons of prostate cancer screening are impacted by any diagnostic or demographic information that will help physicians and their patients at risk for prostate cancer to make more informed decisions about biopsy. In biopsy-positive men, the impact of additional prognostic information regarding treatment options is of equal importance. However, in order to achieve the potential improvement in outcomes, reliable information is needed on the diagnostic accuracy of a new test and its comparators for the outcomes of interest.

Ultimately, direct or indirect evidence is also needed to measure improvement in long-term health outcomes related to the use of the test or risk tools and subsequent decisionmaking. This is particularly true following the recent report of the results of the PIVOT trial. Twelve-year followup of men with localized prostate cancer revealed no difference in all-cause or prostate cancer-specific mortality between men assigned to observation (watchful waiting) versus those randomly assigned to radical prostatectomy treatment.

Limitations of the Comparative Effectiveness Review Process

One limitation of the review process was our decision to craft two separate KQs on biopsy decisionmaking – one for initial biopsies and one for repeat biopsies. We were able to identify only a small number of studies specifically targeting use for these specific populations. A majority of studies included populations with mixed biopsy histories (some patients were candidates for initial and others for repeat biopsy). This was addressed in our analysis, but reflected a characteristic of studies that was not anticipated when we scoped the KQs.

A second limitation was the need to develop an expanded QUADAS framework to address the quality of studies for the three KQs. QUADAS asks if the reference standard results are interpreted without knowledge of the index test results (in this case PCA3). However, it was also important to know if this was also true for the tPSA comparator test, and if partial verification bias was identified. For KQ 3 we observed that the endpoint of interest (patient clinical outcomes) seldom had appropriate clinical information. Again we added an additional QUADAS item to ensure we accounted for this information.

Finally, a procedural limitation was the implementation of the DistillerSR application (and the associated learning curve) concurrently with the beginning of this review. This was balanced by strengths of DistillerSR, such as improved efficiency of abstract review.

Limitations of the Evidence Base

KQs 1 and 2

There were several important limitations of the evidence base identified during the review. First, our inclusion criteria required that PCA3 and the comparator both be measured in the same population. These studies were identified as being "matched." The aim was to reduce the well-known variability in diagnostic test performance due to factors such as underlying population demographics and study entry criteria. By requiring matched comparisons, the analyses would be expected to have reduced heterogeneity, as all differences were computed "within study." However, requiring this matching certainly reduced the number of included studies. It is possible that additional conclusions may have been reached, if unmatched data had been included. On the other hand, the associated increase in variability may have rendered those data unhelpful.

Another challenging aspect of this review was integrating and interpreting the information for diagnostic accuracy. We found data to support the conclusion that PCA3 had slightly higher performance compared with the extent of tPSA elevation. Based on limited data, the two markers seemed to be relatively independent. Better performance would be expected by combining the two. We could have extended our modeling to include both markers, but chose not to because that was beyond the scope of the report. Such modeling was identified as a gap in knowledge.

A weighting scheme could have been developed to address the more precise estimates of effect size associated with larger, multicenter studies compared with smaller studies. Theoretically, weighting could impact the overall strengths of evidence. However, given the uniformly poor quality of individual studies and lack of needed data and analyses (e.g., matched contingency tables), it seemed unlikely that results of the analyses would have been improved by the added complexity of weighting.

There were limited data available when focusing only on initial, or only on repeat, prostate biopsy studies. Several addition studies reported a mixture of initial and repeat biopsies and others did not report this information at all. Examining the most common comparison (difference in AUCs for PCA3 and tPSA) showed no evidence of a relationship between the proportion of initial biopsies and AUC difference. Based on this, we combined all relevant studies, regardless of the proportion of initial biopsies.

The issue of verification bias for tPSA (and related comparators) was raised early in discussion of the analytical framework and Key Questions, and discussed with members of the TEP. Although this bias was not acknowledged by any of the included studies, we still chose to proceed with review of this comparator. An internal modeling exercise was undertaken to investigate the magnitude of the effect, and whether it affected the overall estimate of performance (i.e., ROC curve), performance estimate at a given cutoff (i.e., sensitivity/specificity at a PCA3 cutoff of 35), or both. We concluded that the effect of verification bias, while present, was unlikely to influence the overall estimate of performance (e.g., sensitivity at a given specificity), but had a strong influence on the sensitivity/specificity occurring at a given cutoff. This latter issue was addressed by modeling (Appendix J).

KQ 3

The review of the literature revealed a lack of clinical followup after patients were placed into risk categories defined by the results of PCA3, other biomarker and pathological tests. In 11 of 13 studies, a reference or gold standard clinical endpoint (or validated surrogate) was lacking; in one poor-quality short-term study, PCA3 levels were not associated with disease progression.

Research Gaps

Overview

The PCA3 test is the key marker of this comparative effectiveness review. Its performance has been evaluated against/with six comparators (tPSA, %fPSA, PSA density, PSA velocity and doubling time, complexed PSA and externally validated nomograms) in three clinical scenarios (diagnostic accuracy and other intermediate outcomes in initial and repeat biopsied patients, and long term outcomes). With the exception of analyses of PCA3 and tPSA for the intermediate outcome of diagnostic accuracy, evidence was insufficient to answer the Key Question(s). These then, remain as gaps in evidence. Even for the PCA3/tPSA comparison and diagnostic accuracy, the strength of evidence was low. Thus, virtually all comparisons for all outcomes in this review could be considered gaps in knowledge. As these comparisons have been extensively reviewed in the Results and Discussion, they will not be repeated here.

Instead, the following sections deal with gaps in knowledge and their associated research question/future study design for cross-cutting issues, statistical/methodological issue related to multiple comparisons, or clinical issues that are relevant to multiple comparisons. The first

section focuses on gaps for KQ 1 and KQ 2, while the second section focuses on KQ 3. Within each section, the specific gaps are gathered into two general areas: gaps relating to clinical issues, and those relating to methodological and statistical issues. For each identified gap, a potential study is designed and discussed.

Gaps in Knowledge for KQs 1 and 2

Clinical Gaps in Knowledge

1. Does the addition of PCA3, either alone, or in combination with other markers, change prostate cancer biopsy decisionmaking for the patient or physician? Several studies (and our review) provide evidence that PCA3 may improve individualized risk prediction among men with an initial positive tPSA and/or DRE. However, no information is available on whether the clinical use of PCA3 can be effectively used to change current practice and what educational materials for patients and providers would support this process.
   Future Studies: Researchers might consider routinely combining the tPSA elevations with PCA3 results according to one of several validated nomograms contained in this review. The population could be randomized into those with PCA3 included as part of care (intervention) versus those receiving current care (control). Both groups would have urine samples collected; with the intervention group being immediately tested, while the control group members are tested later. Cutoffs levels in the two groups could be selected separately, such that the sensitivities (proportion of cancers detected) are expected to be similar in both groups. However, fewer biopsies would be expected among the intervention group. This would be expected to result in a higher positive predictive value. Once the study is complete (perhaps one year later), patients could receive additional test results, as if they had been randomized to the other group. This would allow for a fully matched analysis of which cancers were detected by both, as well as comparing the cancer detected by protocol with the other. These data could provide the type of data that would allow PCA3 to become a routine test, or be rejected as a potential contributor to the testing process.
2. What improvement in diagnostic accuracy is needed for any new test (e.g., PCA3) to provide sufficient value to impact biopsy decisionmaking? Were there clear guidance on how much improvement in diagnostic accuracy would be required to impact clinical protocols, the methods required to assess and accept/reject prospective markers would be streamlined. The relative importance of other factors to be considered (e.g., convenience, cost) would also be useful.
   Future Studies: Models can be created for various types of markers (e.g., continuous, categorical) with varying performance characteristics (e.g., low sensitivity but high specificity, high sensitivity but low specificity) that utilize, when possible, existing prostate cancer markers. Health care providers could evaluate the relative improvement in test performance and also provide guidelines regarding other factors that assay developers need to bear in mind.
3. How does PCA3 compare with the two commonly used add on tests of %fPSA and tPSA velocity/doubling time? These comparisons have been singled out, because both comparators have been recommended for clinical implementation (NCCN guidelines) but their use has generated some controversy rather than bringing consensus. Special

attention should be paid toward looking at the relative performance of PCA3 against these two comparators at outcomes of decisionmaking as a way to avoid even further fracturing of protocols based on limited evidence.
Future Studies: Since our restriction to "matched" studies identified no suitable data, it may be necessary to move to a broader set of inclusion criteria that does not require PCA3 and fPSA/tPSA velocity to be measured in the same dataset. However, prior to this undertaking, it would be prudent to validate that such a methodology is likely to provide reliable information (see Methodological and Statistical Gaps in Knowledge, Number 4).

4. Is PCA3 affected by key demographic features known to change risk for prostate cancer (ethnicity, family history)? These features were not well reported in most studies. Their impact on performance of PCA3 (as well as for some of the comparators) is unknown, but may be important.
Future Studies: It may be possible to create collaboration between groups studying PCA3 such that pooled raw data might be created. If funding is available, it may also be possible to collect some of the relevant data that might be missing. From this combined dataset, it might be possible to answer some of these important questions.
5. What is the population from which the convenience samples of biopsied men have been selected? Nearly all of the "matched" studies were convenience samples gathered by centers performing prostate biopsies. These sites should be encouraged to gather information regarding the catchment population as a way to estimate the potential for partial verification bias.
Future Studies: Few, if any studies included in this review identified a well-defined cohort of consecutive men identified as being at risk for prostate cancer and then followed their subsequent screening/diagnostic decision-making. This may require external funding, as a relatively high proportion of men identified as being at high risk will not chose to have a biopsy, according to current guidelines. However, with the collection of data that could reduce partial verification bias, and longer followup, it may be possible to account for, or more clearly delineate the effect of this bias on tPSA and on other related biomarkers.

Methodological and Statistical Gaps in Knowledge

1. What modeling approach/algorithm would allow for the easiest inclusion of new markers while reducing the need for independent verification? Most reported multivariate modeling of prostate cancer risk relies on logistic regression. These models are difficult to compare across studies and do not allow for simple inclusion of new variables without re-computing all coefficients. Other models, such as multivariate overlapping Gaussian distributions, may fit the markers of interest and might allow for easier comparisons as well as the ability to easily add (or subtract) markers as knowledge increases. This could also allow for validation of partial models, if some markers have not been measured.
Future Studies: A review of select literature might provide a model methodology to use in designing an approach to multivariate modeling. One that has been successful in the area of prenatal screening for select congenital abnormalities (e.g., Down syndrome) is based on multi-dimensional overlapping Gaussian distributions. Such testing began in the early 1980s with both a demographic (maternal age) and a second trimester biochemical marker (alpha-fetoprotein). Currently, the most advanced protocol relies on these and three more second trimester and two first trimester biochemical markers, and a first

trimester ultrasound markers. All are successfully combined to produce one risk that has been extensively validated as part of routine screening. Other models to explore might include those used in testing for BRCA1/2 mutations and breast cancer.

2. What factors influence whether partial verification bias impacts the tPSA and/or the matched tPSA/PCA3 ROC curves? Factors that could be explored include the range of cancer rates, the range of verification rates, and the use of continuous versus categorical verification corrections. There have been only a handful of reports on tPSA use that address partial verification bias. A better understanding of this issue is needed if PCA3 is to be properly evaluated in the context of the widespread use of tPSA as triage test for treatment decisions.
   Future Studies: Both modeling and examining select datasets might help provide an answer to this question in the specific setting of interest. Our modeling suggests that verification bias introduced by decision-making relying on tPSA elevations has a strong impact on the sensitivity/specificity at a given tPSA cutoff. However, we found that it does not distort the ROC curve to any great extent. If this can be verified, it might result in the reexamination of previous studies where verification bias was thought to have made the reliable interpretation of results impossible.
3. What absolute cutoffs or continuous values can be assigned to the PCA3 assay across the ROC curve? While the analyses in this review provide an approximate ROC curve allowing interpolation of sensitivity and false positive rates across the range of values, the absolute PCA3 and tPSA cutoffs may not be appropriate in every setting.
   Future Studies: The modeling studies suggested for the previous gap could also help resolve this issue. Is it possible to account for verification bias to such an extent that reliable performance estimates can be computed?
4. Does our review's literature restriction to matched studies provide more consistent and reliable comparisons than had the review used independent summaries of each marker's performance? Given the increasing emphasis on comparative effectiveness analyses, a formal comparison of these two methods might provide useful guidance to future reviews.
   Future Studies: A structured evidence review between the performance of PCA3 and tPSA could be performed without restricting data to matched studies. That is, a summary review of PCA3 performance (via an ROC curve) could be compared with a similar independent summary of the performance of tPSA elevations to identify prostate cancer. The two independent ROC curves could then be compared and the difference compared with our matched differences. As a check on precision, one could examine the tPSA analysis and determine whether the "grey zone" studies differ from the studies having no restriction on the tPSA elevation. Our matched analyses clearly show this difference and are likely an indicator of reliability.
5. Does the reporting of matched analyses improve the usefulness of the dataset? Although our inclusion criteria required PCA3 and a comparator to be measured in the same population, it did not require a formal matched analysis to be reported. Thus, the reports did not allow for a comparison of how many men with cancer were identified by both markers, neither of the markers, or only one or the other marker. Requesting such analyses be performed using existing datasets would help answer this question.
   Future Studies: It may be possible to collaborate with authors of selected studies to obtain original data, in which a true matched analysis can be performed. The study (or

studies) should be large, and have information available about the "aggressiveness" of the cancers identified. Then, two 2x2 tables could be created (one for positive biopsies, one for negative) comparing the results of the two tests at chosen cutoffs. A series of these tables could be created at fixed sensitivities or fixed false positive rates. One could imagine that the data examined this way will provide insights into the types of cancers identified by both, one, or neither of the tests.

6. How can researchers studying PCA3 and other comparators be encouraged to provide proper reporting of statistical details? Proper reporting of statistical information on studies of PCA3 and the comparators was often absent in articles evaluated for this review. These include: confidence intervals, standard errors, prediction limits and other measure of dispersion and precision for all effect measures as well as good summary parameters for their data (e.g., selected centiles, medians, geometric means and trimmed logarithmic standard deviations). All studies identified were of poor quality when rated by QUADAS.
   Future Studies: Statisticians, epidemiologists and others with experience in analyzing screening and diagnostic tests are in the process of setting guidelines for evaluation. Those guidelines should be reviewed with an eye towards ensuring that: (1) appropriate statistics are used (e.g., needed transformations performed), (2) methods to allow for raw data are available for additional analyses, (3) confidence intervals/standard errors are provided more consistently, when possible, and (4) suitable data are provided to allow for joint meta-analyses of multiple tests.
7. How can systematic differences in marker levels due to reagents/manufactures be minimized or accounted for by analysis? Systematic differences between reagents/manufacturers exist for at least some of the markers that can influence the tests performance at fixed mass unit cutoffs.
   Future Studies: Review of external proficiency testing results may provide guidance as to what steps may be needed. Attempting to harmonize results is one possibility, but this can take time and may not ever succeed. Mathematical methods that use a normalizing function (z-score, multiples of the median) could also be explored and tested against existing datasets to see whether more consistent patient-specific information can be provided from the test results.

Gaps in Knowledge for KQ 3

Clinical Gaps in Knowledge

1. What should the gold standard be for defining intermediate outcomes for use in establishing the clinical validity of PCA3? Studies evaluating PCA3 as selection criteria for entering a program of active surveillance have focused on how well PCA3 compares with other selection criteria (tumor volume, tumor grade, clinical stage, Epstein criteria, etc.). These intermediate measures were not well described in most studies and vary considerably between studies.
   Future studies: A consensus conference or statement from a relevant professional organization might help harmonize the definitions of intermediate outcomes such that research studies can be combined. However, the only sure way to determine the appropriateness of intermediate outcomes is to engage in long-term followup of a select

cohort of men with sufficient numbers that even relatively uncommon events (e.g., death due to prostate cancer) can be quantified.

2. How can PCA3 alone or when integrated with one or more comparators be used to improve decisionmaking about whether to choose active surveillance or aggressive treatment for biopsy positive men? No studies have yet examined the impact of PCA3 on decisionmaking when compared with existing criteria such as the Epstein Criteria. There have been no outcomes studies performed to determine how well PCA3 scores predict the behavior of a particular tumor over time.
   Future Studies: This is similar to question 1 (Clinical Gap in Knowledge for KQ 1/KQ 3), but studies would focus on other intermediate outcomes. Based on this review, and the recent FDA approval of PCA3, it should be possible to establish protocols that use PCA3 and begin developing studies that examine and validate intermediate health outcomes.

Methodological and Statistical Gaps in Knowledge

1. Can intermediate outcomes, such as cancer classifications of aggressive or insignificant/ indolent tumors be properly validated? Given that current clinical practice guidelines employ unvalidated, or partially validated intermediate outcomes, it is difficult to design studies that would provide proper validation. Exploration of what study designs or re-analyses of existing dataset might provide stronger validation of select intermediate measures could be undertaken.
   Future Studies: To validate intermediate outcomes, they must be standardized (addressed earlier) and then associated with longer term health outcomes. This might require longitudinal studies with followup of five to ten years. Some of these studies are underway and preliminary results have been reported, but longer timeframes need to be utilized.
2. What is the impact of use of PCA3 on long-term health outcomes when used to help select patients for active surveillance versus aggressive treatment?
   Future Studies: Longitudinal studies that track patients under both active surveillance and aggressive treatment would be helpful in determining this impact. Many of the ongoing studies are focused on only the active surveillance group, and these should be expanded to at least include a subset of those undergoing treatment for comparison.

References

1. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. CA Cancer J Clin. 2010 Sep-Oct;60(5):277-300. PMID: 20610543.
2. Brawley OW. Prostate cancer epidemiology in the United States. World J Urol. 2012 Apr;30(2):195-200. PMID: 22476558.
3. Sutcliffe P, Hummel S, Simpson E, Young T, Rees A, Wilkinson A, Hamdy F, Clarke N, Staffurth J. Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review. Health Technol Assess. 2009 13(1-260. PMID: 19128541.
4. Arcangeli S, Pinzi VArcangeli G. Epidemiology of prostate cancer and treatment remarks. World J Radiol. 2012 Jun 28;4(6):241-6. PMID: 22778875.
5. National Comprehensive Cancer Network Guidelines - Prostate Cancer Early Detection, Version I.2011. www.NCCN.org.
6. Gronberg H. Prostate cancer epidemiology. Lancet. 2003 Mar 8;361(9360):859-64. PMID: 12642065.
7. Freedland SJ. Screening, risk assessment, and the approach to therapy in patients with prostate cancer. Cancer. 2011 Mar 15;117(6):1123-35. PMID: 20960523.
8. Ilic D, O'Connor D, Green S, et al. Screening for prostate cancer: an updated Cochrane systematic review. BJU Int. 2011 107(6):882-91. PMID: 17206534.
9. Yin M, Bastacky S, Chandran U, et al. Prevalence of incidental prostate cancer in the general population: a study of healthy organ donors. J Urol. 2008 Mar;179(3):892-5; discussion 95. PMID: 18207193.
10. SEER Cancer Statistics Review, 1975-2005, 2008. National Cancer Institute, Bethesda, MD. seer.cancer.gov/csr/1975_2005/.
11. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003 95(12):868-78. PMID: 12813170.
12. Basch E, Oliver TK, Vickers A, et al. Screening for Prostate Cancer With Prostate-Specific Antigen Testing: American Society of Clinical Oncology Provisional Clinical Opinion. J Clin Oncol. 2012 Jul 16; PMID: 22802323.
13. Catalona WJ, Smith DS, Ratliff TL, et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med. 1991 Apr 25;324(17):1156-61. PMID: 1707140.
14. Bastian PJ, Carter BH, Bjartell A, et al. Insignificant prostate cancer and active surveillance: from definition to clinical implications. European urology. 2009 Jun;55(6):1321-30. PMID: 19286302.
15. Schroder F, Kattan MW. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. European urology. 2008 Aug;54(2):274-90. PMID: 18511177.
16. Carter HB. Management of low (favourable)-risk prostate cancer. BJU Int. 2011 Dec;108(11):1684-95. PMID: 22077546.
17. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. Cancer causes & control : CCC. 2008 Mar;19(2):175-81. PMID: 18027095.
18. Boyle P. Screening for prostate cancer: have you had your cholesterol measured? BJU Int. 2003 Aug;92(3):191-9. PMID: 12887466.
19. Chou R, Croswell JM, Dana T, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2011 Dec 6;155(11):762-71. PMID: 21984740.
20. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. J Clin Oncol. 2011 Sep 20;29(27):3669-76. PMID: 21825257.

21. Moyer VAon behalf of the USPSTF. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2012 Jul 17;157(2):120-34. PMID: 22801674.

22. Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. CA Cancer J Clin. 2011 60(70-98. PMID:

23. Prostate-Specific Antigen Best Practice Statement: 2009 Update. American Urological Association. www.auanet.org/content/media/psa09.pdf.

24. Rosenthal SASandler HM. Treatment strategies for high-risk locally advanced prostate cancer. Nat Rev Urol. 2010 Jan;7(1):31-8. PMID: 20062072.

25. Thompson IM, Ankerst DP, Chi C, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. J Natl Cancer Inst. 2006 Apr 19;98(8):529-34. PMID: 16622122.

26. Amling CL, Catalona WJ, Klein EA. Deciding whom to biopsy. Urol Oncol. 2010 Sep-Oct;28(5):542-5. PMID: 20816613.

27. Simmons MN, Berglund RKJones JS. A practical guide to prostate cancer diagnosis and management. Cleve Clin J Med. 2011 78(5):321-31. PMID: 21536828.

28. Djavan B, Waldert M, Zlotta A, et al. Safety and morbidity of first and repeat transrectal ultrasound guided prostate needle biopsies: results of a prospective European prostate cancer detection study. J Urol. 2001 Sep;166(3):856-60. PMID: 11490233.

29. D' Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clincally localized prostate cancer. JAMA. 1998 Sep 16;280(11):969-74. PMID: 9749478.

30. Epstein JI, Walsh PC, Carmichael M, et al. Pathologic and clinical findings to predict tumor extert of nonpalpable (stage T1c) prostate cancer. Jama. 1994 271(368-74. PMID:

31. Perdona S, Cavadas V, Di Lorenzo G, et al. Prostate cancer detection in the "grey area" of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. Eur Urol. 2011 Jan;59(1):81-7. PMID: 20947244.

32. Dahabreh IJ, Chung M, Balk EM, et al. Active surveillance in men with localized prostate cancer: a systematic review. Ann Intern Med. 2012 Apr 17;156(8):582-90. PMID: 22351515.

33. Cheng L, Montironi R, Bostwick DG, et al. Staging of prostate cancer. Histopathology. 2012 Jan;60(1):87-117. PMID: 22212080.

34. Chun FK, de la Taille A, van Poppel H, et al. Prostate cancer gene 3 (PCA3): development and internal validation of a novel biopsy nomogram. Eur Urol. 2009 Oct;56(4):659-67. PMID: 19304372.

35. Eastham JA, Scardino PTKattan MW. Predicting an optimal outcome after radical prostatectomy: the trifecta nomogram. J Urol. 2008 Jun;179(6):2207-10; discussion 10-1. PMID: 18423693.

36. Huang Y, Isharwal S, Haese A, et al. Prediction of patient-specific risk and percentile cohort risk of pathological stage outcome using continuous prostate-specific antigen measurement, clinical stage and biopsy Gleason score. BJU Int. 2011 May;107(10):1562-9. PMID: 20875091.

37. Boorjian SA, Karnes RJ, Viterbo R, et al. Long-term survival after radical prostatectomy versus external-beam radiotherapy for patients with high-risk prostate cancer. Cancer. 2011 Jul 1;117(13):2883-91. PMID: 21692049.

38. Hinnen KA, Roeloffzen EM, Battermann JJ, et al. Survival after prostate brachytherapy in patients aged 60 years and younger. BJU Int. 2011 June;107(12):1906-11. PMID: 21062393.

39. Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. 1999 Dec 1;59(23):5975-9. PMID: 10606244.

40. Durand X, Moutereau S, Xylinas E, et al. Progensa PCA3 test for prostate cancer. Expert Rev Mol Diagn. 2011 Mar;11(2):137-44. PMID: 21405964.

41. Makarov DV, Loeb S, Getzenberg RH, et al. Biomarkers for prostate cancer. Ann Rev Medicine. 2009 60(139-51. PMID: 18947298.

42. de Kok JB, Verhaegh GW, Roelofs RW, et al. DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. Cancer Res. 2002 May 1;62(9):2695-8. PMID: 11980670.

43. Meng FJ, Shan A, Jin L, et al. The expression of a variant prostate-specific antigen in human prostate. Cancer Epidemiol Biomarkers Prev. 2002 Mar;11(3):305-9. PMID: 11895882.

44. Day JR, Jost M, Reynolds MA, et al. PCA3: from basic molecular science to the clinical lab. Cancer Lett. 2011 Feb 1;301(1):1-6. PMID: 21093148.

45. de la Taille A. Progensa PCA3 test for prostate cancer detection. Expert Rev Mol Diagn. 2007 Sep;7(5):491-7. PMID: 17892357.

46. Hessels D, Klein Gunnewiek JM, van Oort I, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol. 2003 Jul;44(1):8-15; discussion 15-6. PMID: 12814669.

47. Groskopf J, Aubin SM, Deras IL, et al. APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. Clin Chem. 2006 Jun;52(6):1089-95. PMID: 16627561.

48. Sokoll LJ, Ellis W, Lange P, et al. A multicenter evaluation of the PCA3 molecular urine test: pre-analytical effects, analytical performance, and diagnostic accuracy. Clin Chim Acta. 2008 Mar;389(1-2):1-6. PMID: 18061575.

49. Thompson IM, Ankerst DP, Chi C, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. JAMA. 2005 Jul 6;294(1):66-70. PMID: 15998892.

50. Oesterling JE, Jacobsen SJ, Chute CG, et al. Serum prostate-specific antigen in a community-based population of healthy men. Establishment of age-specific reference ranges. JAMA. 1993 Aug 18;270(7):860-4. PMID: 7688054.

51. Sokoll LJ, Witte DL, Klee GG, et al. Redesigned proficiency testing materials improve survey outcomes for prostate-specific antigen. Arch Pathol Lab Med. 2000 124(1608-13. PMID:

52. Schreiber WE, Endres DB, McDowell GA, et al. Comparison of fresh frozen serum to proficiency testing material in College of American Pathologists surveys: alpha-fetoprotein, carcinoembryonic antigen, human chorionic gonadotropin, and prostate-specific antigen. Arch Pathol Lab Med. 2005 129(331-337): PMID:

53. Garg UC, Howanitz JH, Nakamura RM, et al. Production, analysis, and characterization of reference materials for prostate-specific antigen. Arch Pathol Lab Med. 1995 119(1104-08. PMID:

54. Klee GG, Schryver PG, Kisabith RM. Analytic bias specifications based on the analysis of effects on performance of medical guidelines. Scand J Clin Lab Invest. 1999 59(509-12. PMID:

55. Westgard JO, Westgard SA. An assessment of metrics for analytic quality using σ performance data from proficiency testing surveys and the CLIA criteria for acceptable performance. Am J Clin Pathol. 2006 125(343-54. PMID:

56. Tosoian J, Loeb S. PSA and beyond: the past, present, and future of investigative biomarkers for prostate cancer. Sci World J. 2010 10(1919-31. PMID: 20890581.

57. Roddam AW, Duffy MJ, Hamdy FC, et al. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/ml: systematic review and meta-analysis. Eur Urol. 2005 Sep;48(3):386-99; discussion 98-9. PMID: 15982797.

58. Lee R, Localio AR, Armstrong K, et al. A meta-analysis of the performance characteristics of the free prostate-specific antigen test. Urology. 2006 Apr;67(4):762-8. PMID: 16600352.

59. Benson MC, Whang IS, Pantuck A, et al. Prostate specific antigen densivity; a means of distinguishing benigh prostatic hypertrophy and prostate cancer. J Urol. 1992 147(815-16. PMID:

60. Giannarini G, Scott CA, Moro U, et al. Are PSA density and PSA density of the transition zone more accurate than PSA in predicting the pathological stage of clinically localized prostate cancer? Urol Oncol. 2008 Jul-Aug;26(4):353-60 PMID: 18367100.

61. Kundu SD, Roehl KA, Yu X, et al. Prostate specific antigen density correlates with features of prostate cancer aggressiveness. J Urol. 2007 Feb;177(2):505-09. PMID: 17222621.

62. Thanigasalam R, Mancuso P, Tsao K, et al. Prostate-specific antigen velocity (PSAV): a practical role for PSA? ANZ journal of surgery. 2009 Oct;79(10):703-6. PMID: 19878164.

63. Vickers AJ, Savage C, O'Brien MF, et al. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. J Clin Oncol. 2009 Jan 20;27(3):398-403. PMID: 19064972.

64. Vickers AJBrewster SF. PSA Velocity and Doubling Time in Diagnosis and Prognosis of Prostate Cancer. Br J Med Surg Urol. 2012 Jul 1;5(4):162-68. PMID: 22712027.

65. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008 Oct;180(4):1303-8; discussion 08. PMID: 18707724.

66. Epstein JI, Chan DW, Sokoll LJ, et al. Nonpalpable stage T1c prostate cancer: prediction of insignificant disease using free/total prostate specific antigen levels and needle biopsy findings. J Urol. 1998 Dec;160(6 Pt 2):2407-11. PMID: 9817393.

67. Ploussard G, Epstein JI, Montironi R, et al. The contemporary concept of significant versus insignificant prostate cancer. Eur Urol. 2011 Aug;60(2):291-303. PMID: 21601982.

68. D'Amico AV, Cote K, Loffredo M, et al. Determinants of prostate cancer-specific survival after radiation therapy for patients with clinically localized prostate cancer. J Clin Oncol. 2002 Dec 1;20(23):4567-73. PMID: 12454114.

69. Schroder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med. 2012 Mar 15;366(11):981-90. PMID: 22417251.

70. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; August 2011. AHRQ Publication No. 10(11)-EHC063-EF. www.effectivehealthcare.ahrq.gov /index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=318

71. Methods Guide for Medical Test Reviews. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. effectivehealthcare.ahrq.gov/ehc/products/246/558/Methods-Guide-for-Medical-Test-Reviews_Full-Guide_20120530.pdf

72. Jüni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol. 2002 Feb 1;31(1):115-23. PMID: 11914306.

73. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009 Oct;62(10):1013-20. PMID: 19230606.

74. U.S. Preventive Services Task Force Procedure Manual. Rockville, MD: Agency for Healthcare Research and Quality, July 2008. AHRQ Publication No. 08-051180-EF.

75. Higgins JPT, Deeks JJ, Altman DG on behalf of the Cochrane Statistical Methods Group. Special topics in statistics. In: Cochrane Handbook for Systematic Reviews of Interventions. J. P. T. Higgins and S. Green (Eds). West Sussex, England:Wiley-Blackwell, pp. 481-529, 2010.

76. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006 6(Mar 6):529-36. PMID: 16519814.

77. Begg CBGreenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. . Biometrics. 1983 Mar;39(1):207-13. PMID: 6871349.

78. Punglia RS, D'Amico AV, Catalona WJ, et al. Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. N Engl J Med. 2003 Jul 24;349(4):335-42. PMID: 12878740.

79. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--Agency for Healthcare Research and Quality and the Effective Health-care Program. J Clin Epidemiol. 2010 May;63(5):513-23. PMID: 19595577.

80. Ruiz-Aragon J, Marquez-Pelaez S. [Assessment of the PCA3 test for prostate cancer diagnosis: a systematic review and meta-analysis]. Actas urologicas espanolas. 2010 Apr;34(4):346-55. PMID: 20470697.

81. Auprich M, Haese A, Walz J, et al. External validation of urinary PCA3-based nomograms to individually predict prostate biopsy outcome. Eur Urol. 2010 Nov;58(5):727-32. PMID: 20619529.

82. Haese A, de la Taille A, van Poppel H, et al. Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. Eur Urol. 2008 Nov;54(5):1081-8. PMID: 18602209.

83. Marks LS, Fradet Y, Deras IL, et al. PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. Urology. 2007 Mar;69(3):532-5. PMID: 17382159.

84. van Gils MP, Hessels D, van Hooij O, et al. The time-resolved fluorescence-based PCA3 test on urinary sediments after digital rectal examination; a Dutch multicenter validation of the diagnostic performance. Clin Cancer Res. 2007 Feb 1;13(3):939-43. PMID: 17289888.

85. Fradet Y, Saad F, Aprikian A, et al. uPM3, a new molecular urine test for the detection of prostate cancer. Urology. 2004 Aug;64(2):311-5; discussion 15-6. PMID: 15302485.

86. Tinzl M, Marberger M, Horvath S, et al. DD3PCA3 RNA analysis in urine--a new perspective for detecting prostate cancer. Eur Uroly. 2004 Aug;46(2):182-6; discussion 87. PMID: 15245811.

87. van Gils MP, Hessels D, Hulsbergen-van de Kaa CA, et al. Detailed analysis of histopathological parameters in radical prostatectomy specimens and PCA3 urine test results. Prostate. 2008 Aug 1;68(11):1215-22. PMID: 18500693.

88. Kusuda Y, Miyake H, Kurahashi T, et al. Assessment of optimal target genes for detecting micrometastases in pelvic lymph nodes in patients with prostate cancer undergoing radical prostatectomy by real-time reverse transcriptase-polymerase chain reaction. Urol Oncol. 2011 May 18; PMID: 21600799.

89. Tosoian JJ, Loeb S, Kettermann A, et al. Accuracy of PCA3 measurement in predicting short-term biopsy progression in an active surveillance program. J Urol. 2010 Feb;183(2):534-8. PMID: 20006883.

90. Adam A, Engelbrecht MJ, Bornman MS, et al. The role of the PCA3 assay in predicting prostate biopsy outcome in a South African setting. BJU Intl. 2011 Apr 20; PMID: 21507188.

91. Aubin SM, Reid J, Sarno MJ, et al. PCA3 molecular urine test for predicting repeat prostate biopsy outcome in populations at risk: validation in the placebo arm of the dutasteride REDUCE trial. J Urol. 2010 Nov;184(5):1947-52. PMID: 20850153.

92. Auprich M, Augustin H, Budaus L, et al. A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. BJU Int. 2012 Jun;109(11):1627-35. PMID: 21939492.

93. Bollito E, De Luca S, Cicilano M, et al. Prostate cancer gene 3 urine assay cutoff in diagnosis of prostate cancer: a validation study on an Italian patient population undergoing first and repeat biopsy. Anal Quant Cytol Histol. 2012 Apr;34(2):96-104. PMID: 22611765.

94. Cao DL, Ye DW, Zhang HL, et al. A multiplex model of combining gene-based, protein-based, and metabolite-based with positive and negative markers in urine for the early diagnosis of prostate cancer. Prostate. 2011 May 15;71(7):700-10. PMID: 20957673.

95. de la Taille A, Irani J, Graefen M, et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. J Urol. 2011 Jun;185(6):2119-25. PMID: 21496856.

96. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol. 2008 Apr;179(4):1587-92. PMID: 18295257.

97. U.S. Food and Drug Administration. Summary of Safety and Effectiveness Data: PROGENSA PCA3 Assay, 2012. www.accessdata.fda.gov/cdrh_docs/pdf10/P100033b.pdf.

98. Ferro M, Bruzzese D, Perdona S, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta. 2012 Aug 16;413(15-16):1274-8. PMID: 22542564.

99. Goode RR, Marshall SJ, Duff M, et al. Use of PCA3 in detecting prostate cancer in initial and repeat prostate biopsy patients. Prostate. 2012 May 14; PMID: 22585386.

100. Hessels D, van Gils MP, van Hooij O, et al. Predictive value of PCA3 in urinary sediments in determining clinico-pathological characteristics of prostate cancer. Prostate. 2010 Jan 1;70(1):10-6. PMID: 19708043.

101. Mearini E, Antognelli C, Del Buono C, et al. The combination of urine DD3(PCA3) mRNA and PSA mRNA as molecular markers of prostate cancer. Biomarkers. 2009 Jun;14(4):235-43. PMID: 19489685.

102. Nyberg M, Ulmert D, Lindgren A, et al. PCA3 as a diagnostic marker for prostate cancer: a validation study on a Swedish patient population. Scand J Urol Nephrol. 2010 Dec;44(6):378-83. PMID: 20961267.

103. Ochiai A, Okihara K, Kamoi K, et al. Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. Int J Urol. 2011 Mar;18(3):200-5. PMID: 21332814.

104. Ouyang B, Bracken B, Burke B, et al. A duplex quantitative polymerase chain reaction assay based on quantification of alpha-methylacyl-CoA racemase transcripts and prostate cancer antigen 3 in urine sediments improved diagnostic accuracy for prostate cancer. J Urol. 2009 Jun;181(6):2508-13; discussion 13-4. PMID: 19371911.

105. Pepe PAragona F. PCA3 score vs PSA free/total accuracy in prostate cancer diagnosis at repeat saturation biopsy. Anticancer Res. 2011 Dec;31(12):4445-9. PMID: 22199313.

106. Ploussard G, Haese A, Van Poppel H, et al. The prostate cancer gene 3 (PCA3) urine test in men with previous negative biopsies: does free-to-total prostate-specific antigen ratio influence the performance of the PCA3 score in predicting positive biopsies? BJU Int. 2010 Oct;106(8):1143-7. PMID: 20230386.

107. Rigau M, Morote J, Mir MC, et al. PSGR and PCA3 as biomarkers for the detection of prostate cancer in urine. Prostate. 2010 Dec 1;70(16):1760-7. PMID: 20672322.

108. Roobol MJ, Schroder FH, van Leeuwen P, et al. Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: exploring the value of PCA3 for a first-line diagnostic test. European urology. 2010 Oct;58(6):475-81. PMID: 20637539.

109. Schilling D, Hennenlotter J, Munz M, et al. Interpretation of the prostate cancer gene 3 in reference to the individual clinical background: implications for daily practice. Urol Int. 2010 85(2):159-65. PMID: 20424427.

110. Wang R, Chinnaiyan AM, Dunn RL, et al. Rational approach to implementation of prostate cancer antigen 3 into clinical care. Cancer. 2009 Sep 1;115(17):3879-86. PMID: 19517474.

111. Wu AK, Reese AC, Cooperberg MR, et al. Utility of PCA3 in patients undergoing repeat biopsy for prostate cancer. Prostate Cancer Prostatic Dis. 2012 Mar;15(1):100-5. PMID: 22042252.

112. Auprich M, Chun FK, Ward JF, et al. Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. Euro Urol. 2011 Jan;59(1):96-105. PMID: 20980098.

113. Durand X, Xylinas E, Radulescu C, et al. The value of urinary prostate cancer gene 3 (PCA3) scores in predicting pathological features at radical prostatectomy. BJU Int. 2012 Jul;110(1):43-49. PMID: 22221521.

114. Liss MA, Santos R, Osann K, et al. PCA3 molecular urine assay for prostate cancer: association with pathologic features and impact of collection protocols. World J Urol. 2011 Oct;29(5):683-8. PMID: 21152924.

115. Nakanishi H, Groskopf J, Fritsche HA, et al. PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. J Urol. 2008 May;179(5):1804-9; discussion 09-10. PMID: 18353398.

116. Ploussard G, Durand X, Xylinas E, et al. Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance. European urology. 2011 Mar;59(3):422-9. PMID: 21156337.

117. van Poppel H, Haese A, Graefen M, et al. The relationship between Prostate CAncer gene 3 (PCA3) and prostate cancer significance. BJU Int. 2012 Feb;109(3):360-6. PMID: 21883822.

118. Vlaeminck-Guillem V, Devonec M, Colombel M, et al. Urinary PCA3 score predicts prostate cancer multifocality. J Urol. 2011 Apr;185(4):1234-9. PMID: 21334023.

119. Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. J Urol. 2008 Nov;180(5):1975-8; discussion 78-9. PMID: 18801539.

120. Roddam AW, Hamdy FC, Allen NE, et al. The impact of reducing the prostate-specific antigen threshold and including isoform reflex tests on the performance characteristics of a prostate-cancer detection programme. BJU Int. 2007 Sep;100(3):514-7. PMID: 17542987.

121. Otto SJ, Moss SM, Maattanen L, et al. PSA levels and cancer detection rate by centre in the European Randomized Study of Screening for Prostate Cancer. Eur J Cancer. 2010 Nov;46(17):3053-60. PMID: 21047586.

122. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 2005 5(17. PMID: 15840177.

123. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. N Engl J Med. 2012 Jul 19;367(3):203-13. PMID: 22808955.

124. Epstein JI, Walsh PC, Carmichael M, et al. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. JAMA. 1994 Feb 2;271(5):368-74. PMID: 7506797.

125. Boorjian SA, Karnes RJ, Rangel LJ, et al. Mayo Clinic validation of the D'amico risk group classification for predicting survival following radical prostatectomy. J Urol. 2008 Apr;179(4):1354-60; discussion 60-1. PMID: 18289596.

126. Jeldres C, Suardi N, Walz J, et al. Validation of the contemporary epstein criteria for insignificant prostate cancer in European men. Eur Urol. 2008 Dec;54(6):1306-13. PMID: 18083294.

127. Oon SF, Watson RW, O'Leary JJ, et al. Epstein criteria for insignificant prostate cancer. BJU Int. 2011 Aug;108(4):518-25. PMID: 21320276.

128. Laxman B, Morris DS, Yu J, et al. A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. Cancer Res. 2008 Feb 1;68(3):645-9. PMID: 18245462.

129. Shappell SB, Fulmer J, Arguello D, et al. PCA3 urine mRNA testing for prostate carcinoma: patterns of use by community urologists and assay performance in reference laboratory setting. Urology. 2009 Feb;73(2):363-8. PMID: 18995890.

130. Wang Y, Sun G, Pan JG, et al. Performance of tPSA and f/tPSA for prostate cancer in Chinese. A systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2006 9(4):374-8. PMID: 16926855.

Abbreviations and Acronyms

ACMG	American College of Medical Genetics
ACS	American Cancer Society
AHRQ	Agency for Healthcare Research and Quality
AMP	Association for Molecular Pathology
AS	active surveillance
AUC	area under the receiver operating characteristics (ROC) curve
AUA	American Urological Association
ASAP	atypical small acinar proliferation
ASCO	American Society of Clinical Oncology
BPH	benign prostatic hyperplasia
CAP	College of American Pathologists
Cc	cubic centimeter (measure of volume)
CER	comparative effectiveness review
CI	confidence interval
CU	clinical utility
CV	clinical validity OR coefficient of variation
DA	diagnostic accuracy
DD3	differential display code 3
DOR	diagnostic odds ratio
DRE	digital rectal examination
ECE	extracapsular extension
EPICOT	evidence, population, intervention, comparison, outcome, timestamp
ERSPC	European Randomized Study of Screening for Prostate Cancer
EVN	externally validated nomogram
FDA	Food and Drug Administration
FPR	false positive rate
GRADE	grading of recommendations assessment, development and evaluation
GS	Gleason score
IQR	Inter-quartile range
LR	likelihood ratio
Mg	milligram
Ml	milliliter
mRNA	messenger RNA (ribonucleic acid)
Ng	nanogram
NICE	UK National Institute for Health and Clinical Excellence
NIH	National Institutes of Health
NCCN	National Cancer Consortium Network
NCI	National Cancer Institute
NPV	negative predictive value
NR	not reported
OAPR	odds of being affected given a positive result
OANR	odds of being affected given a negative result
OR	odds ratio
PCa	prostate cancer, prostatic cancer
PCA3	prostate cancer antigen 3

PCPT	Prostate Cancer Prevention Trial
PCRI	Prostate Cancer Research Institute
PICOTS	patient, intervention, comparator, outcome, timing, and setting
PIN	prostatic intraepithelial neoplasia
PPV	positive predictive value
PSA	prostate-specific antigen
cPSA	complexed prostate-specific antigen
%fPSA	free prostate-specific antigen
tPSA	total prostate-specific antigen
PSAD	prostate-specific antigen density
PSADT	prostate-specific antigen doubling time
PSAV	prostate-specific antigen velocity
RCT	randomized controlled trial
RP	radical prostatectomy
RT-PCR	real-time – polymerase chain reaction
SD	standard deviation
SER	systematic evidence review
TEP	technical expert panel
TMA	transcription-mediated amplification
TURP	transurethral resection of the prostate
TV	tumor volume
U.S.	United States

Appendix A. Search Strategies for PCA3 Testing for the Diagnosis and Management of Prostate Cancer

02/09/11

((prostate cancer antigen 3, human [SUBSTANCE NAME] OR pca3 [TIAB] OR (prostate cancer antigen 3 [TIAB]) OR (prostatic cancer antigen 3 [TIAB]) OR DD3 antigen, human [SUBSTANCE NAME] OR (differential display code 3 [TIAB]) OR dd3 [TIAB]) OR (prostatic neoplasms [MH] AND ((clinical* [TIAB] AND (significan* [TIAB] OR importan* [TIAB])) OR aggressive [TIAB] OR biops* [TIAB]) AND (nomogram [TIAB] OR (neural [TIAB] AND network [TIAB])))) OR ((((((clinical* [TIAB] AND (significan* [TIAB] OR importan* [TIAB])) OR aggressive [TIAB] OR biops* [TIAB]) AND prostate-specific antigen [MH]) AND prostatic neoplasms [MH]) AND (predict* [TIAB] OR prognos* [TIAB])))

Revised Searches (PubMed only)

"prostate cancer antigen 3, human" [Supplementary Concept] OR
("differential display code 3 antigen" OR DD3) Field: Title/Abstract OR
(PCA3 OR "prostate cancer antigen 3") Field: Title/Abstract

- This was the test-specific set= 208 in PubMed

Additionally -
"Prostatic Neoplasms"[Mesh] OR "prostatic neoplasms" OR "prostate neoplasms" OR "prostatic cancer" OR "prostate cancer"
AND
(nomogram OR (neural AND network) OR antigen OR antigens) Field: Title/Abstract
AND
((clinical* AND (significan* OR importan* OR aggressive OR biops*)) OR predict* OR prognos* OR (select* OR decid* OR decision* OR choos* OR choice*)) Field: Title/Abstract
NOT the test-specific set
AND Limits: Humans, Publication Date from 1/1/1990 to search date.

PUBMED on 8/9/11

Test specific search
"prostate cancer antigen 3, human" [Supplementary Concept] OR PCA3 OR DD3 OR DD3PCA3 OR "DD3(PCA3)" OR "prostate cancer gene 3" OR "prostate cancer antigen 3" OR progensa OR ("differential display code 3" AND (prostate OR prostatic))
AND
Limits: Humans, English = 159

Comparators search
1. PSA/total PSA
"Prostate-Specific Antigen"[Mesh] OR "total PSA" OR "total prostate specific antigen" OR "prostate specific antigen" OR (PSA AND (prostate OR prostatic))

AND
Publication type: meta-analysis OR Subset: systematic review OR "meta-analysis" OR metaanalysis OR "systematic review"
AND
Limits: Humans, English = 480

2. PSA velocity/percent free PSA/complexed PSA/externally validated nomograms
"PSA velocity" OR "prostate specific antigen velocity" OR "percent free PSA" OR "free prostate specific antigen" OR "complexed PSA" OR "c-PSA" OR "complexed prostate specific antigen" OR (nomogram* AND (prostatic OR prostate))
AND
Publication Types: Clinical Trial, Meta-Analysis, Practice Guideline, Randomized Controlled Trial, Clinical Trial, Phase II, Clinical Trial, Phase III, Clinical Trial, Phase IV, Comparative Study, Controlled Clinical Trial, Multicenter Study OR Subset: systematic review
AND
Limits: Humans, English = 521

EMBASE on 8/15/11

Test specific search
'prostate cancer antigen 3, human' OR pca3 OR dd3 OR dd3pca3 OR 'dd3(pca3)' OR 'prostate cancer gene 3' OR 'prostate cancer antigen 3' OR progensa OR ('differential display code 3' AND ('prostate'/exp OR prostatic))
AND
‘prostate'/exp OR prostatic
AND
Limits: Humans, English = 64

Comparators search
1. PSA/Total PSA
'total psa' OR 'total prostate specific antigen' OR 'prostate specific antigen'/exp OR (psa AND ('prostate'/exp OR prostatic))
AND
'meta analysis'/exp OR 'systematic review'/exp OR 'metaanalysis'/exp
AND
Limits: Humans, English 258

2. PSA velocity/percent free PSA/complexed PSA/externally validated nomograms
'psa velocity' OR 'prostate specific antigen velocity' OR 'percent free PSA' OR 'free prostate specific antigen' OR 'complexed psa' OR 'c-psa' OR 'complexed prostate specific antigen' OR (nomogram* AND (prostatic OR 'prostate'/exp))
AND
'meta analysis'/exp OR 'systematic review'/exp OR 'metaanalysis' OR 'randomized clinical trial' OR 'randomised clinical trial' OR 'comparative trial' OR 'controlled trial'/exp OR random OR 'comparison'/exp
AND
'major clinical study'/de

AND
Limits: Humans, English =125

Cochrane Central
Test names searched for anything that was not in the other two databases' results = 2 new records

Additional comparator searching in Cochrane = 160 new records

Appendix A Addendum

Literature Search Update – 05/15/2012

A full update of the comprehensive search initially conducted on August 9, 2012 was not performed. Rather, the search was focused on PCA3, with the aim of identifying any subsequently published studies that could provide additional relevant data. Searches were conducted searches in PubMed®, EMBASE.COM, The Cochrane Library, and the Clinicaltrials.gov (NCT) database. The results included 40 unique citations for published articles not captured in the initial search, as well as 43 abstracts from scientific meetings (e.g., American Society for Clinical Oncology, Journal of Urology Annual Conference) and updated information on three relevant ongoing clinical trials.

Search Strategies for Updating of Evidence for PCA3 Testing for the Diagnosis and Management of Prostate Cancer

PUBMED on 5/15/2012
("prostate cancer antigen 3, human" [Supplementary Concept] OR "PCA3" OR "DD3" OR "DD3PCA3" OR "DD3(PCA3)" OR "prostate cancer gene 3" OR "prostate cancer antigen 3" OR "progensa" OR "differential display code 3") AND ("prostate" OR "prostatic")

Limits: English, Dates 8/1/2011 to 5/15/2012
Results: 31 citations

EMBASE.COM on 5/15/2012
'prostate cancer antigen 3, human' OR pca3 OR dd3 OR dd3pca3 OR 'dd3(pca3)' OR 'prostate cancer gene 3' OR 'prostate cancer antigen 3' OR progensa OR 'differential display code 3' AND ('prostate'/exp OR prostatic)

Limits: English, Dates 8/1/2011 to 5/15/2012
Results: 76 citations

Cochrane Central on 5/15/2012
prostate cancer antigen 3 OR pca3 OR dd3 OR dd3pca3 OR dd3(pca3) OR prostate cancer gene 3 OR prostate cancer antigen 3 OR progensa OR differential display code 3' AND ('prostate'/exp OR prostatic)

Results: No new trials identified

Appendix B. Data Elements

- Study description and design, including:
  - Country, institutions and enrollment period
  - Enrollment number and flow of subjects through PCA3 and comparator testing, prostate biopsy, treatment and followup
  - Source of funding and authors' disclosures of industry relationship(s)
  - Blinding of index and comparator test results to pathologists and of biopsy/prostatectomy results to laboratorians conducting tests
- Participant characteristics, including
  - Demographics of the study population
  - Criteria for study inclusion (e.g., age, race, elevated tPSA, abnormal DRE, previous negative or positive biopsy, family history)
  - Comorbidities or potential effect modifiers
- Prostate biopsy (KQ1-3) and radical prostatectomy (KQ3) findings, including:
  - Cores per biopsy, positive cores per biopsy
  - Gleason scores
  - Other biomarkers (e.g., PSA density)
  - Pathological markers
  - Percentage of 'insignificant findings' based on identified criteria
  - Clinical and pathological staging of tumor from prostatectomy
- PCA3 specimens and assay characteristics, including:
  - Method of collection
  - Handling/storage
  - PCA3 assay used (e.g., specific test or method, housekeeping gene used, reporting unit)
- PCA3 and comparator test results, including:
  - Specified comparators were total PSA, PSA velocity or doubling time, percent free PSA, PSA density, complexed PSA, externally validated nomograms or risk assessment programs
  - Cutoffs/thresholds/action points
  - Summary measures (e.g., mean or median values), stratified by negative or positive biopsy result, cutoff, or other variables
- Intermediate outcomes, including:
  - Diagnostic accuracy data, including area under the ROC (receiver operating characteristics) curve (AUC), diagnostic odds ratios, clinical sensitivity and specificity, positive and negative predictive values
  - Data on decisionmaking related to biopsy, with definition of study design, description of participants and instruments used, and outcome measures
  - Data on harms related to biopsy, with definition of study design, description of participants and instruments used, and findings
- Long-term outcomes, including:
  - Mortality, including overall and prostate cancer-specific mortality and 10-year survival

    - Morbidity, including local progression, distant metastases, pain, and biochemical failure
    - Treatment-related morbidity, including urinary incontinence, impotence, rectal incontinence and prostatitis
    - Quality-of-life measures
- Statistical analyses, including:
    - Statistical tests used
    - Confidence intervals for performance estimates
    - p values for comparisons
    - Assessment of potential biases
- Quality assessment:
    - Selection of participants to avoid bias
    - Adequate descriptions of study design and process and reasons for cases lost
    - Use of blinding
    - Methods are described below in the section entitle Assessment of Methodological Quality of Individual Studies

Appendix C. Search Strategy for Grey Literature

Regulatory Information
FDA
Source: http://www.fda.gov/default.htm
Date searched: 5/15/2012
Search strategy: key word “PCA3”
Records: 219

Clinical trial registries
NIH database
Source: http://clinicaltrials.gov/
Date searched: 5/15/2012
Search strategy: PCA3 [ALL-FIELDS] AND "Completed" [OVERALL-STATUS]
Records: 6

BioMed central
Source: http://www.controlled-trials.com/mrct/
Date searched: 8/15/2011
Search strategy: “PCA3” for completed trials
Records: 0

WHO International Clinical Trials Registry Platform Search Portal
Source: http://apps.who.int/trialsearch/
Date searched: 8/15/2011
Search strategy: Search String = “PCA3” for ALL recruitment status trials
Records: 0

Conference papers and abstracts
American College of Medical Genetics (ACMG)
Source: http://submissions.miracd.com/acmg/
Date searched: 8/15/2011
Search strategy: search string “PCA3 OR Prostate Cancer OR Prostate Cancer Screening”
Records: 2

American Society of Clinical Oncology (ASCO)
Source: http://www.gucasym.org/PastSymposia.aspx
Date searched: 8/15/2011
Search strategy: search string “PCA3”
Records: 6

American Urological Association (AUA)
Source: http://www.auanet.org/content/clinical-practice-guidelines/clinical-practice-guidelines.cfm
Date searched: 8/15/2011
Search strategy: search string “PCA3”

Records: 26

Association of Molecular Pathology (AMP)
Source: http://www.amp.org/meetings/past_meetings.cfm
Date searched: 8/15/2011
Search strategy: not searchable without login
Records: 0

College of American Pathologists (CAP)
Source: http://www.cap.org/apps/cap.portal?_nfpb=true&_pageLabel=reference
Date searched: 8/15/2011
Search strategy: search string “PCA3 OR Prostate Cancer OR Prostate Cancer Screening”
Records: 240

National Comprehensive Cancer Network (NCCN) Annual Congress
Source: http://www.nccn.org/index.asp
Date searched: 8/15/2011
Search strategy: search string “PCA3”
Records: 0

Prostate Cancer Research Initiative (PCRI)
Source: http://www.prostate-cancer.org/pcricms/
Date searched: 8/15/2011
Search strategy: search string “PCA3”
Records: 0

Organizations publishing Guidance or Review Documents
Agence d’évaluation des technologies et des modes d’intervention en santé (AETMIS)
Source: http://www.aetmis.gouv.qc.ca/site/home.phtml
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
Records: 0

Australia and New Zealand Horizon Scanning Network (ANZHSN)
Source: http://www.horizonscanning.gov.au/
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
Records: 0

Canadian Agency for Drugs and Technologies in Health (CADTH)
Source: http://cadth.ca/
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").

Records: 3 (all excluded; not relevant)

The Cochrane Collaboration (Cochrane)
Source: http://summaries.cochrane.org/
Date searched: 8/15/2011
Search strategy: search string string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR"differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
Records: 0

European network for Health Technology Assessment (EUnetHTA)
Source: http://www.eunethta.eu/Public/Search/
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
Records: 0

EuroGenTest
Source: http://www.eurogentest.org/
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
Records: 0

National Guideline Clearinghouse (NGC)
Source: http://www.guideline.gov/
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
– searches done in different combinations using the above combinations.
Records: 4

National Institute for Clinical Excellence (NICE)
Source: http://www.nice.org.uk/
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
Records: 1 (excluded; not relevant)

United Kingdom Health Technology Assessment Programme (UK HTA)
Source: http://www.hta.ac.uk/
Date searched: 8/15/2011
Search strategy: search string ("PCA3" OR "prostate cancer antigen 3" OR "DD3" OR "differential display code 3" OR "biomarkers") AND ("prostate cancer" OR "prostatic cancer").
Records: 7

Government documents

RePORTER

Source: http://projectreporter.nih.gov/reporter.cfm
Date searched: 8/15/2011
Search strategy: key word “PCA3”
Records: 26

HSRPROJ
Source: http://wwwcf.nlm.nih.gov/hsr_project/home_proj.cfm
Date searched: 8/15/2011
Search strategy: key word “PCA3”
Records: 0

AHRQ GOLD
Source: http://gold.ahrq.gov/projectsearch/
Date searched: 8/15/2011
Search strategy: key word “PCA3”
Records: 0

Manufacturer database
Source: GenProbe Response to the Request for Scientific Information
Date posted: 12/7/2011
Date searched: Not applicable
Search strategy: Not applicable
Records: 27

References for Review and Abstraction from the Grey Literature Search

PUBMED ID: NA

Authors: Submitted by Gen-Probe, Inc.

Title: Summary of Safety and Effectiveness Data: PROGENSA PCA3 Assay.

Citation: U.S. Food and Drug Administration, http://www.accessdata.fda.gov/cdrh_docs/pdf10/P100033b.pdf, accessed June 1, 2012.

Selected Test Manufacturers for Contact

Manufacturer	Registered® / Trademark Name™	Test Type	Comment
GenProbe	Gen-Probe PROGENSA PCA3	FDA Approval announced 2/17/12	Responded pre-FDA approval
Quest Diagnostics	PCA3 Diagnostic Test	LDT	Did not respond
Bostwick Laboratories	PCA3*Plus*®	LDT	Did not respond
Laboratory Corporation of America	CaPDetect: PCA3	LDT	Did not respond

Contact Information

Gen Probe Incorporated	10210 Genetic Center Drive San Deigo, CA 92121 Phone: 858-410-8000 http://www.genprobe.com
Quest Diagnostics	3 Giralda Farms Madison, NJ 07940 Phone: 800-222-0446 http://www.questdiagnostics.com
Bostwick Laboratories	4355 Innslake Drive Glen Allen, VA 23060 **Phone:** 877-865-3262 http://www.bostwicklaboratories.com
Laboratory Corporation of America	358 South Main Street Burlington, NC 27215 Phone: 336-584-5171 http://www.labcorp.com

Appendix D. Articles Excluded at Full-Text Level

Reason for Exclusion: Study participants did not meet study population inclusion criteria **AND** invalid study design

1. S. Abuzallouf, I. Dayes and H. Lukka 2004. Baseline staging of newly diagnosed prostate cancer: A summary of the literature Journal of Urology, 171(6 I): 2122-2127.
2. S. Agrawal and W. D. Dunsmuir 2009. Molecular markers in prostate cancer. Part I: predicting lethality Asian J Androl, 11(1): 14-21.
3. G. L. Andriole, Jr. 2010. Screening for prostate cancer BMJ, 341: c4538.
4. D. C. Aziz and R. B. Barathur 1993. Prostate-specific antigen and prostate volume: a meta-analysis of prostate cancer screening criteria J Clin Lab Anal, 7(5): 283-92.
5. K. Belej, O. Kaplan, O. Kohler, J. Kocarek and P. Fojtik 2010. Prostate cancer gene 3 (PCA3) in prognosis after robotic assisted radical prostatectomy - Initial experience European Urology, Supplements, 9(6): 630.
6. A. Bjartell 2007. PSA and Prostate Cancer Screening: The Challenge of the New Millennium European Urology, 52(5): 1284-1286.
7. D. G. Bostwick, V. E. Gould, J. Qian, M. Susani and M. Marberger 2006. Prostate cancer detected by uPM3: radical prostatectomy findings Mod Pathol, 19(5): 630-3.
8. W. J. Catalona, A. W. Partin, M. G. Sanda, J. T. Wei, G. G. Klee, C. H. Bangma, K. M. Slawin, L. S. Marks, S. Loeb, D. L. Broyles, S. S. Shin, A. B. Cruz, D. W. Chan, L. J. Sokoll, W. L. Roberts, R. H. van Schaik and I. A. Mizrahi 2011. A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range J Urol, 185(5): 1650-5.
9. J. S. Chung, H. Y. Choi, H. R. Song, S. S. Byun, S. Seo, C. Song, J. S. Cho, S. E. Lee, H. Ahn, E. S. Lee, W. J. Kim, M. K. Chung, T. Y. Jung, H. S. Yu and Y. D. Choi 2010. Preoperative nomograms for predicting extracapsular extension in Korean men with localized prostate cancer: a multi-institutional clinicopathologic study J Korean Med Sci, 25(10): 1443-8.
10. D. Connolly, R. Hutton and P. F. Keane 2011. Re: Monique J. Roobol, Fritz H. Schroder, Pim van Leeuwen, et al. Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: exploring the value of PCA3 for a first-line diagnostic test. Eur Urol 2010;58:475-81 Eur Urol, 59(3): e9; author reply e10-1.
11. E. D. Crawford, P. F. Pinsky, D. Chia, B. S. Kramer, R. M. Fagerstrom, G. Andriole, D. Reding, E. P. Gelmann, D. L. Levin and J. K. Gohagan 2006. Prostate specific antigen changes as related to the initial prostate specific antigen: data from the prostate, lung, colorectal and ovarian cancer screening trial. J Urol, 175: 1286-90; discussion 1290.
12. E. D. Crawford and P. A. Abrahamsson 2008. PSA-based screening for prostate cancer: how does it compare with other cancer screening tests? Eur Urol, 54(2): 262-73.
13. A. V. D'Amico and M. H. Chen 2009. Pretreatment prostate-specific antigen velocity and the risk of death from prostate cancer in the individual with low-risk prostate cancer J Clin Oncol, 27(22): 3575-6.
14. M. Ding, X. Cao, H. Xu, et al. Prostate cancer-specific and potent antitumor effect of a DD3-controlled oncolytic virus harboring the PTEN gene. PloS One. 2012 7(4): e35153 Epub 2012 Apr 11.
15. T. Dorff and S. Tucker 2009. Prostate cancer in younger men poses clinical and research challenges Community Oncology, 6(9): 427-430.
16. X. Durand, S. Moutereau, E. Xylinas and A. de la Taille 2011. Progensa PCA3 test for prostate cancer Expert Rev Mol Diagn, 11(2): 137-44.
17. X. Durand, E. Xylinas, C. Radulescu, et al. The value of urinary prostate cancer gene 3 (PCA3) scores in predicting pathological features at radical prostatectomy. BJU Int. 2012 Jul;110(1):43-9.
18. D. U. Ekwueme, L. A. Stroud and Y. Chen 2007. Cost analysis of screening for, diagnosing, and staging prostate cancer based on a systematic review of published studies Prev Chronic Dis, 4(4): A100.
19. V. Ficarra, G. Novara and F. Zattoni 2010. The role of the prostate cancer antigen 3 (PCA3)

test for the diagnosis of prostate cancer in the era of opportunistic prostate-specific antigen screening Eur Urol, 58(4): 482-4; discussion 484-5.
20. J. Fichtner 2000. The management of prostate cancer in patients with a rising prostate-specific antigen level BJU International, 86(2): 181-190.
21. S. Fontenete, J. Silva, A.L. Teixeira , et al. Controversies in using urine samples for Prostate Cancer detection: PSA and PCA3 expression analysis. International Braz J Urol. 2011 Nov-Dec;37(6):719-26.
22. S. Fontenete, A. Nogueira, F. Pina, et al. Molecular study of the PCA3 gene: genotypic analysis of PCA3 polymorphism -845G>A and metastatic prostate cancer. Genetic testing and molecular biomarkers. 2012 May;16(5):418-22.
23. F. Galasso, R. Giannella, P. Bruni, R. Giulivo, V. R. Barbini, V. Disanto, R. Leonardi, V. Pansadoro and G. Sepe 2010. PCA3: a new tool to diagnose prostate cancer (PCa) and a guidance in biopsy decisions. Preliminary report of the UrOP study Arch Ital Urol Androl, 82(1): 5-9.
24. S. R. Goyal, V. H. Talib and S. K. Khurana 1999. An overview of PSA/percent free PSA with special reference to recent trends in diagnosis of prostatic cancer Indian J Pathol Microbiol, 42(2): 171-8.
25. E. P. Gregorio, J. P. Grando, E. E. Saqueti, S. H. Almeida, H. A. Moreira and M. A. Rodrigues 2007. Comparison between PSA density, percent free PSA percentage and PSA density in the transition zone in the detection of prostate cancer in patients with serum PSA between 4 and 10 ng/mL Int Braz J Urol, 33(2): 151-60.
26. A. J. Grillo-Lopez 2005. The ODAC chronicles: Part 5. Prostate cancer endpoints Expert Review of Anticancer Therapy, 5(3): 405-410.
27. R. L. Grubb, 3rd and G. L. Andriole 2006. Can preoperative PSA doubling time and PSA velocity predict outcomes following radical prostatectomy? Nat Clin Pract Urol, 3(6): 306-7.
28. K. H. Gulkesen, I. T. Koksal, U. Bilge and O. Saka 2010. Comparison of methods for prediction of prostate cancer in Turkish men with PSA levels of 0-10 ng/mL J BUON, 15(3): 537-42.
29. M. Haid, D. Rabin, K. M. King, C. M. Feinstein, K. L. Janson, S. R. Levine, D. L. Mutchnik, E. A. Lambiase and R. Bradley 1994. Digital rectal examination, serum prostate specific antigen, and prostatic ultrasound: how effective is this diagnostic triad? J Surg Oncol, 56(1): 32-8.
30. M.R. Haythorn, R.J. Ablin. Prostate-specific antigen testing across the spectrum of prostate cancer. Biomarkers in medicine. 2011 5(4):515-26.
31. A. J. Henderson, K. R. Ghani, J. Cook, M. Fahey, J. Schalken and R. Thilagarajah 2010. The role of PCA3 testing in patients with a raised prostate-specific antigen level after Greenlight photoselective vaporization of the prostate J Endourol, 24(11): 1821-4.
32. T. R. Herrmann, A. S. Merseburger and M. Burchardt 2010. Prostate cancer: novel aspects of diagnostics and surgical technology World J Urol, 28(6): 665.
33. A. Horwich 2004. Prostate cancer management Annals of Oncology, 15(SUPPL. 4): iv307-iv312.
34. D. Ilic, D. O'Connor, S. Green and T. Wilt 2006. Screening for prostate cancer Cochrane Database Syst Rev, 3: CD004720.
35. D. G. Ingram and M. W. Kattan 2010. Risk grouping versus risk continuum in patients with clinically localized prostate cancer: a taxometric test J Urol, 184(5): 1937-41.
36. F. H. Jansen, M. Roobol, G. Jenster, F. H. Schroder and C. H. Bangma 2009. Screening for prostate cancer in 2008 II: the importance of molecular subforms of prostate-specific antigen and tissue kallikreins Eur Urol, 55(3): 563-74.
37. Y. Kakehi 2011. Re: Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance Eur Urol, 60(1): 178-9.
38. P. I. Karakiewicz and G. C. Hutterer 2007. Predicting outcomes in patients with urologic cancers Curr Opin Support Palliat Care, 1(3): 153-68.
39. M. W. Kattan and P. T. Scardino 2002. Prediction of progression: nomograms of clinical utility Clin Prostate Cancer, 1(2): 90-6.
40. M. S. Katz, J. A. Efstathiou, A. V. D'Amico, M. W. Kattan, M. G. Sanda, P. L. Nguyen, M. R. Smith, P. R. Carroll and A. L. Zietman 2010. The 'CaP Calculator': an online decision support tool for clinically localized prostate cancer BJU Int, 105(10): 1417-22.
41. R. Kirby 2007. PCA3 improves diagnosis of prostate cancer Practitioner, 251(1690): 18, 21, 23.

42. R. Kirby, J.M. Fitzpatrick. Optimising repeat prostate biopsy decisions and procedures. BJU Int. 2012 Jun;109(12):1750-4.
43. S. Langley, H.U. Ahmed, B. Al-Qaisieh, et al. Report of a consensus meeting on focal low dose rate brachytherapy for prostate cancer. BJU Int. 2012 109(SUPPL. 1):7-16.
44. M. F. Lavin, R. Clarke and R. A. Gardiner 2009. Differential expression of PCA3 and BMCC1 in prostate cancer Prostate, 69(16): 1713-4; author reply 1715.
45. L. Laskiewicz, Z. Jiang, D.C. Altieri, et al. The search for a better prostate cancer biomarker. Journal of Urology. 2011 186(5):1758-59.
46. M. Law 2004. Screening without evidence of efficacy British Medical Journal, 328(7435): 301-302.
47. G. L. Lee, A. Dobi and S. Srivastava 2011. Prostate cancer: diagnostic performance of the PCA3 urine test Nat Rev Urol, 8(3): 123-4.
48. S. Loeb 2008. Does PCA3 help identify clinically significant prostate cancer? Eur Urol, 54(5): 980-1.
49. S. Loeb and A. W. Partin 2010. PCA3 Urinary Biomarker for Prostate Cancer Rev Urol, 12(4): e205-6.
50. S. Loeb 2009. Prostate cancer: is PSA velocity useful? Nat Rev Urol, 6(6): 305-6.
51. V. Lorusso 2002. Prostate carcinoma Tumori, 88(SUPPL. 1): S125-S127.
52. D. V. Makarov, S. Loeb, R. H. Getzenberg and A. W. Partin 2009. Biomarkers for prostate cancer Annu Rev Med, 60: 139-51.
53. R. M. Martin, D. Gunnell, F. Hamdy, D. Neal, A. Lane and J. Donovan 2006. Continuing controversy over monitoring men with localized prostate cancer: a systematic review of programs in the prostate specific antigen era J Urol, 176(2): 439-49.
54. A. J. Martin, C. D. Cheli, K. Sterling, M. Ward, S. Pollard, D. Lifsey, D. Mercante, L. Martin and W. Rayford 2006. Prostate specific antigen isoforms and human glandular kallikrein 2 - Which offers the best screening performance in a predominantly black population? Journal of Urology, 175(1): 104-107.
55. C. H. Martinez, V. Chalasani and J. Chin 2009. Molecular biomarkers in prostate cancer Expert Opinion on Medical Diagnostics, 3(4): 345-353.
56. J. B. Nelson, A. R. Allen, S. M. Hulting, J. D. Isaacson and D. S. Sleep 2004. Prostate-specific antigen doubling time as a predictor of prostate cancer disease progression und survival. Abstract [Journal unknown] 22: 394.
57. L. Ng, N. Karunasinghe, C.S. Benjamin, et al. Beyond PSA: Are new prostate cancer biomarkers of potential value to New Zealand doctors? New Zealand Medical Journal. 2012 125(1353):59-86.
58. J.C. Nickel, M. Speakman. Should we really consider Gleason 6 prostate cancer? BJU Int. 2012 109(5):645-46.
59. L. Nogueira, R. Corradi and J. A. Eastham 2010. Other biomarkers for detecting prostate cancer BJU Int, 105(2): 166-9.
60. N. Oakley 1998. Clinical implications of prostate-specific antigen (PSA) Current Opinion in Urology, 8(5): 401-406.
61. M. Oliveira, V. Marques, A. P. Carvalho and A. Santos 2011. Head-to-head comparison of two online nomograms for prostate biopsy outcome prediction BJU International, 107(11): 1780-1783.
62. P. Pepe, F. Aragona. Does an inflammatory pattern at primary biopsy suggest a lower risk for prostate cancer at repeated saturation prostate biopsy? Urol Int. 2011 87(2):171-74.
63. A. Pelekanos, J. Beardi, D. Jonas and G. M. Oremek 2008. Isoforms of PSA (prostate-specific antigen) in the diagnosis of prostate cancer Clinical and Experimental Medical Letters, 49(1): 23-25.
64. P. Puppo 2007. Repeated Negative Prostate Biopsies with Persistently Elevated or Rising PSA: A Modern Urologic Dilemma European Urology, 52(3): 639-641.
65. G. K. Reddy and T. B. Gibson 2005. Identification of potential therapeutic targets using microarray data in prostate cancer: A large-scale metaanalysis by oncomine Clinical Prostate Cancer, 3(4): 209-210.
66. O. Riesterer, L. Milas, K.K. Ang 2007. Use of molecular biomarkers for predicting the response to radiotherapy with or without chemotherapy. J Clin Oncol, 25(26): 4075-83.
67. M. Reni and A. Bolognesi 1998. Prognostic value of prostate specific antigen before, during and after radiotherapy. Review. Cancer Treat Rev, 24: 91-99.
68. C.D. Roberson, S. Atay, C. Gercel-Taylor, et al. Tumor-derived exosomes as mediators of disease and potential diagnostic biomarkers. Cancer Biomarkers. 2010 8(4-5):281-87.
69. M. J. Roobol, F. H. Schroder, G. L. van Leenders, D. Hessels, R. C. van den Bergh, T. Wolters and

P. J. van Leeuwen 2010. Performance of prostate cancer antigen 3 (PCA3) and prostate-specific antigen in Prescreened men: reproducibility and detection characteristics for prostate cancer patients with high PCA3 scores (>/= 100) Eur Urol, 58(6): 893-9.
70. M. J. Roobol, S. Carlsson and J. Hugosson 2011. Meta-analysis finds screening for prostate cancer with PSA does not reduce prostate cancer-related or all-cause mortality but results likely due to heterogeneity - the two highest quality studies identified do find prostate cancer-related mortality reductions Evid Based Med, 16(1): 20-1.
71. M. J. Roobol 2011. Contemporary role of prostate cancer gene 3 in the management of prostate cancer Curr Opin Urol, 21(3): 225-9.
72. P. L. Ross, P. T. Scardino and M. W. Kattan 2001. A catalog of prostate cancer nomograms J Urol, 165(5): 1562-8.
73. M. Salagierski and J. A. Schalken 2010. PCA3 and TMPRSS2-ERG: Promising biomarkers in prostate cancer diagnosis Cancers, 2(3): 1432-1440.
74. G. Sardana, B. Dowell and E. P. Diamandis 2008. Emerging biomarkers for the diagnosis and prognosis of prostate cancer Clin Chem, 54(12): 1951-60.
75. N. Satake, M. Ohori, C. Yu, M. W. Kattan, Y. Ohno, A. Miyakawa, T. Hatano and M. Tachibana 2010. Development and internal validation of a nomogram predicting extracapsular extension in radical prostatectomy specimens Int J Urol, 17(3): 267-72.
76. J. A. Schalken 2009. Towards Early and More Specific Diagnosis of Prostate Cancer? Beyond PSA: New Biomarkers Ready for Prime Time European Urology, Supplements, 8(3): 97-102.
77. P. Schellhammer 2003. Clinical trials in prostate cancer BJU International, 92(3): 186-187.
78. D. Schilling, T. de Reijke, B. Tombal, A. de la Taille, J. Hennenlotter and A. Stenzl 2009. The Prostate Cancer gene 3 assay: indications for use in clinical practice BJU Int, 105(4): 452-5.
79. F. H. Schroder, L. J. Denis and M. Roobol 2003. The story of the European Randomized Study of Screening for Prostate Cancer BJU International, Supplement, 92(2): 1-13.
80. J.A. Sioss, R. Bhiladvala, W. Pan, et al. Nanoresonator chip-based RNA sensor strategy for detection of circulating tumor cells: response using PCA3 as a prostate cancer marker. Nanomedicine: Nanotechnology, Biology, and Medicine. Epub 2011 Nov 22.
81. D. P. Smith, E. Banks, M. S. Clements, R. A. Gardiner and B. K. Armstrong 2009. Evidence-based uncertainty: recent trial results on prostate-specific antigen testing and prostate cancer mortality Med J Aust, 191(4): 199-200.
82. C. Stephan, H. Rittenhouse, H. Cammann, M. Lein, M. Schrader, S. Deger, K. Miller and K. Jung 2009. New markers and multivariate models for prostate cancer detection Anticancer Res, 29(7): 2589-600.
83. P. Sutcliffe, S. Hummel, E. Simpson, T. Young, A. Rees, A. Wilkinson, F. Hamdy, N. Clarke and J. Staffurth 2009. Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review Health Technol Assess, 13(5): iii, xi-xiii 1-219.
84. R. Thanigasalam, P. Mancuso, K. Tsao and P. Rashid 2009. Prostate-specific antigen velocity (PSAV): a practical role for PSA? ANZ J Surg, 79(10): 703-6.
85. J. Tosoian and S. Loeb 2010. PSA and beyond: the past, present, and future of investigative biomarkers for prostate cancer ScientificWorldJournal, 10: 1919-31.
86. G. Tzimagiorgis, E.Z. Michailidou, A. Kritis, et al. Recovering circulating extracellular or cell-free RNA from bodily fluids. Cancer Epidemiology. 2011 35(6):580-89.
87. D. Ulmert, A. M. Serio, M. F. O'Brien, C. Becker, J. A. Eastham, P. T. Scardino, T. Bjork, G. Berglund, A. J. Vickers and H. Lilja 2008. Long-term prediction of prostate cancer: prostate-specific antigen (PSA) velocity is predictive but does not improve the predictive accuracy of a single PSA measurement 15 years or more before cancer diagnosis in a large, representative, unscreened population J Clin Oncol, 26(6): 835-41.
88. R. C. van den Bergh, S. Roemeling, M. J. Roobol, T. Wolters, F. H. Schroder and C. H. Bangma 2008. Prostate-specific antigen kinetics in clinical decision-making during active surveillance for early prostate cancer--a review Eur Urol, 54(3): 505-16.
89. J. Vickers, C. Savage, M. F. O'Brien and H. Lilja 2009. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer J Clin Oncol, 27(3): 398-403.
90. V. Vlaeminck-Guillem, M. Bandel, M. Cottancin, et al. Chronic prostatitis does not influence

urinary PCA3 score. Prostate. 2012 Apr;72(5):549-54.
91. D. E. Whittemore, E. J. Hick, M. R. Carter, J. W. Moul, A. J. Miranda-Sousa and W. J. Sexton 2008. Significance of tertiary Gleason pattern 5 in Gleason score 7 radical prostatectomy specimens J Urol, 179(2): 516-22; discussion 522.
92. A. L. Zietman 1995. Time to second prostate specific antigen (PSA) failure is a surrogate endpoint for prostate cancer death in prospective trials of therapy for localized disease. Abstract no: 1013, [Journal unknown]32: 229.

Reason for Exclusion: Study participants did not meet study population inclusion criteria

1. S. M. Aubin, J. Reid, M. J. Sarno, A. Blase, J. Aussie, H. Rittenhouse, R. S. Rittmaster, G. L. Andriole and J. Groskopf 2011. Prostate Cancer Gene 3 Score Predicts Prostate Biopsy Outcome in Men Receiving Dutasteride for Prevention of Prostate Cancer: Results From the REDUCE Trial Urology, 78(2): 380-5.
2. B. Laxman, D.S. Morris, J. Yu, et al. 2008. A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. Cancer research, 68:645-9.
3. F. Saad 2005. UPM3: review of a new molecular diagnostic urine test for prostate cancer. The Canadian jounal of urology;12 Suppl 1:40-43; discussion 99-100.
4. S.S. Salami, F. Schmidt, B. Laxman, et al. Combining urinary detection of TMPRSS2:ERG and PCA3 with serum PSA to predict diagnosis of prostate cancer. Urologic oncology May 19, 2011. Web posting prior to publication.
5. V. N. Talesa, C. Antognelli, C. Del Buono, F. Stracci, M. R. Serva, E. Cottini and E. Mearini 2009. Diagnostic potential in prostate cancer of a panel of urinary molecular tumor markers Cancer Biomark, 5(6): 241-51.

Reason for Exclusion: Invalid Study Design

1. A. Benchikh, C. Savage, A. Cronin, G. Salama, A. Villers, H. Lilja and A. Vickers 2010. A panel of kallikrein markers can predict outcome of prostate biopsy following clinical work-up: an independent validation study from the European Randomized Study of Prostate Cancer screening, France BMC Cancer, 10: 635.
2. L. Benecchi, A. M. Pieri, C. Destro Pastizzaro and M. Potenzoni 2008. Optimal measure of PSA kinetics to identify prostate cancer Urology, 71(3): 390-4.
3. A. P. Berger, M. Deibl, A. Strasak, J. Bektic, A. E. Pelzer, H. Klocker, H. Steiner, G. Fritsche, G. Bartsch and W. Horninger 2007. Large-scale study of clinical impact of PSA velocity: long-term PSA kinetics as method of differentiating men with from those without prostate cancer Urology, 69(1): 134-8.
4. U. Capitanio, A. Briganti, A. Gallina, N. Suardi, P. I. Karakiewicz, F. Montorsi and V. Scattoni 2010. Predictive models before and after radical prostatectomy Prostate, 70(12): 1371-8.
5. F. K. Chun, M. Graefen, A. Briganti, A. Gallina, J. Hopp, M. W. Kattan, H. Huland and P. I. Karakiewicz 2007. Initial biopsy outcome prediction--head-to-head comparison of a logistic regression-based nomogram versus artificial neural network Eur Urol, 51(5): 1236-40; discussion 1241-3.
6. J. S. Chung, H. Y. Choi, H. R. Song, S. S. Byun, S. I. Seo, C. Song, J. S. Cho, S. E. Lee, H. Ahn, E. S. Lee, T. K. Hwang, W. J. Kim, M. K. Chung, T. Y. Jung, H. S. Yu and Y. D. Choi 2011. Nomogram to predict insignificant prostate cancer at radical prostatectomy in Korean men: a multi-center study Yonsei Med J, 52(1): 74-80.
7. A. V. D'Amico, M. Hui-Chen, A. A. Renshaw, B. Sussman, K. A. Roehl and W. J. Catalona 2006. Identifying men diagnosed with clinically localized prostate cancer who are at high risk for death from prostate cancer J Urol, 176(6 Pt 2): S11-5.
8. S. Y. Eskicorapci, L. Turkeri, E. Karabulut, C. Cal, H. Akpinar, S. Baltaci, K. Baykal, M. W. Kattan and H. Ozen 2009. Validation of two preoperative Kattan nomograms predicting recurrence after radical prostatectomy for localized prostate cancer in Turkey: a

multicenter study of the Uro-oncology Society Urology, 74(6): 1289-95.
9. K. J. Gancarczyk, H. Wu, D. G. McLeod, C. Kane, L. Kusuda, R. Lance, J. Herring, J. Foley, D. Baldwin, J. T. Bishoff, D. Soderdahl and J. W. Moul 2003. Using the percentage of biopsy cores positive for cancer, pretreatment PSA, and highest biopsy Gleason sum to predict pathologic stage after radical prostatectomy: the Center for Prostate Disease Research nomograms Urology, 61(3): 589-95.
10. A. A. Haroun, A. S. Hadidy, Z. M. Awwad, C. F. Nimri, W. S. Mahafza and E. S. Tarawneh 2011. Utility of free prostate specific antigen serum level and its related parameters in the diagnosis of prostate cancer Saudi J Kidney Dis Transpl, 22(2): 291-7.
11. A. A. Haroun 2011. New indicator for prostate gland biopsy when malignancy is in question Saudi J Kidney Dis Transpl, 22(1): 61-6.
12. D. Hessels, J. M. Klein Gunnewiek, I. van Oort, H. F. Karthaus, G. J. van Leenders, B. van Balken, L. A. Kiemeney, J. A. Witjes and J. A. Schalken 2003. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer Eur Urol, 44(1): 8-15; discussion 15-6.
13. R. M. Hoffman, D. L. Clanon, M. Chavez and J. C. Peirce 2002. Using multiple cutpoints for the free-to-total prostate specific antigen ratio improves the accuracy of prostate cancer detection Prostate, 52(2): 150-8.
14. R. M. Hoffman, D. L. Clanon, B. Littenberg, J. J. Frank and J. C. Peirce 2000. Using the free-to-total prostate-specific antigen ratio to detect prostate cancer in men with nonspecific elevations of prostate-specific antigen levels J Gen Intern Med, 15(10): 739-48.
15. A. Kefi, B. Irer, I. Ozdemir, B. Tuna, Y. Goktay, K. Yorukoglu and A. Esen 2005. Predictive value of the international prostate symptom score for positive prostate needle biopsy in the low-intermediate prostate-specific antigen range Urol Int, 75(3): 222-6.
16. R. S. Kirby, J. M. Fitzpatrick and J. Irani 2009. Prostate cancer diagnosis in the new millennium: strengths and weaknesses of prostate-specific antigen and the discovery and clinical evaluation of prostate cancer gene 3 (PCA3) BJU Int, 103(4): 441-5.
17. T. Klatte, M. Waldert, M. de Martino, et al. Age-specific PCA3 score reference values for diagnosis of prostate cancer. World J Urol. 2012 Jun;30(3):405-10.
18. T. Kobayashi, T. Kawahara, K. Nishizawa, K. Ogura, K. Mitsumori and Y. Ide 2005. Value of prostate volume measurement using transabdominal ultrasonography for the improvement of prostate-speci fi c antigen-based cancer detection Int J Urol, 12(10): 881-5.
19. R. Lee, A. R. Localio, K. Armstrong, S. B. Malkowicz and J. S. Schwartz 2006. A meta-analysis of the performance characteristics of the free prostate-specific antigen test Urology, 67(4): 762-8.
20. S. Loeb, K. A. Roehl, C. S. Thaxton and W. J. Catalona 2008. Combined prostate-specific antigen density and biopsy features to predict "clinically insignificant" prostate cancer Urology, 72(1): 143-7.
21. S. Loeb, A. Kettermann, L. Ferrucci, P. Landis, E. J. Metter and H. B. Carter 2008. PSA doubling time versus PSA velocity to predict high-risk prostate cancer: data from the Baltimore Longitudinal Study of Aging Eur Urol, 54(5):1073-80.
22. G. Lughezzani, A. Briganti, P. I. Karakiewicz, M. W. Kattan, F. Montorsi, S. F. Shariat and A. J. Vickers 2010. Predictive and prognostic models in radical prostatectomy candidates: a critical analysis of the literature Eur Urol, 58(5): 687-700.
23. G. Lughezzani, L. Budaus, H. Isbarn, M. Sun, P. Perrotte, A. Haese, F. K. Chun, T. Schlomm, T. Steuber, H. Heinzer, H. Huland, F. Montorsi, M. Graefen and P. I. Karakiewicz 2010. Head-to-head comparison of the three most commonly used preoperative models for prediction of biochemical recurrence after radical prostatectomy Eur Urol, 57(4): 562-8.
24. M. C. Miller, G. J. O'Dowd, A. W. Partin and R. W. Veltri 2001. Contemporary use of complexed PSA and calculated percent percent free PSA for early detection of prostate cancer: Impact of changing disease demographics Urology, 57(6): 1105-1111.
25. A. M. Moreira, J. C. Presti, Jr., W. J. Aronson, M. K. Terris, C. J. Kane, C. L. Amling, L. L. Sun, J. W. Moul and S. J. Freedland 2010. Postoperative prostate-specific antigen nadir improves accuracy for predicting biochemical recurrence after radical prostatectomy: Results from the Shared Equal Access Regional Cancer Hospital (SEARCH) and Duke Prostate Center databases Int J Urol, 17(11): 914-22.

26. N. Mutlu, L. N. Turkeri, F. Yencilek, A. Demir and K. Emerk 2009. Complexed prostate specific antigen: better test in the diagnosis of prostate cancer for the clinically relevant 2.5-4 ng/ml total PSA range Can J Urol, 16(2): 4558-67.
27. K. Okihara, O. Ukimura, T. Nakamura, Y. Mizutani, A. Kawauchi, Y. Naya, M. Uchida, T. Ogiwara and T. Miki 2004. Can complexed prostate specific antigen enhance prostate cancer detection in Japanese men? European Urology, 46(1): 57-64.
28. T. Okegawa, M. Kinjo, M. Ohta, I. Miura, S. Horie, K. Nutahara and E. Higashihara 2003. Predictors of prostate cancer on repeat prostatic biopsy in men with serum total prostate-specific antigen between 4.1 and 10 ng/mL Int J Urol, 10(4): 201-6.
29. D. J. Parekh, D. P. Ankerst, D. Troyer, S. Srivastava and I. M. Thompson 2007. Biomarkers for prostate cancer detection J Urol, 178(6): 2252-9.
30. H. K. Park, K. Y. Lee, K. H. Kim, H. Jung, S. J. Yoon and T. B. Kim 2010. Intermediate versus low or high prostate-specific antigen density level: comparison of cancer detection rate between 12- and 18-core prostate biopsy Scand J Urol Nephrol, 44(6): 391-8.
31. J. K. Parsons, M. K. Brawer, C. D. Cheli, A. W. Partin and R. Djavan 2004. Complexed prostate specific antigen (PSA) reduces unnecessary prostate biopsies in the 2.6-4.0 ng/mL range of total PSA BJU Int, 94(1): 47-50.
32. J.R. Prensner, M.A. Rubin, J.T. Wei, et al. Beyond PSA: The next generation of prostate cancer biomarkers. Science Translational Medicine. 2012 4(127):127rv3.
33. A. W. Roddam, M. J. Duffy, F. C. Hamdy, A. M. Ward, J. Patnick, C. P. Price, J. Rimmer, C. Sturgeon, P. White and N. E. Allen 2005. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/ml: systematic review and meta-analysis Eur Urol, 48(3): 386-99; discussion 398-9.
34. M. J. Roobol, I. W. van der Cruijsen and F. H. Schroder 2004. No reason for immediate repeat sextant biopsy after negative initial sextant biopsy in men with PSA level of 4.0 ng/mL or greater (ERSPC, Rotterdam). Urology, 63(5): 892-7; discussion 897-9.
35. J. Rubio-Briones, A. Fernández-Serra, L. Ramírez, L. Rubio, A. Collado, J. Casanova, A. Gómez-Ferrer, J.v. Ricós, J.L. Monrós, R. Dumont, B. Ortiz, I. Iborra, Z. García-Casado, E. Solsona, J.A. López-Guerrero. Actas Urol Esp. 2011 Nov-Dec;35(10):589-96. Epub 2011 Jun 22. Spanish / English translation.
36. J. Ruiz-Aragon and S. Marquez-Pelaez 2010. Assessment of the PCA3 test for prostate cancer diagnosis: a systematic review and meta-analysis. Actas Urol Esp [translated into English by the journal] 34:346-55.
37. I. F. San Francisco, L. Werner, M. M. Regan, M. B. Garnick, G. Bubley and W. C. DeWolf 2011. Risk stratification and validation of prostate specific antigen density as independent predictor of progression in men with low risk prostate cancer during active surveillance J Urol, 185(2): 471-6.
38. A. Schroder and M. W. Kattan 2008. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review Eur Urol, 54(2): 274-90.
39. A. Sciarra, V. Panebianco, S. Cattarino, et al. Multiparametric magnetic resonance imaging of the prostate can improve the predictive value of the urinary prostate cancer antigen 3 test in patients with elevated prostate-specific antigen levels and a previous negative biopsy. BJU Int. Epub 2012 May 4.
40. S. B. Shappell, J. Fulmer, D. Arguello, B. S. Wright, J. R. Oppenheimer and M. J. Putzi 2009. PCA3 urine mRNA testing for prostate carcinoma: patterns of use by community urologists and assay performance in reference laboratory setting Urology, 73(2): 363-8.
41. S. F. Shariat, P. I. Karakiewicz, N. Suardi and M. W. Kattan 2008. Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature Clin Cancer Res, 14(14): 4400-7.
42. S. F. Shariat, P. I. Karakiewicz, V. Margulis and M. W. Kattan 2008. Inventory of prostate cancer predictive tools Curr Opin Urol, 18(3): 279-96.
43. J. M. Song, C. B. Kim, H. C. Chung and R. L. Kane 2005. Prostate-specific antigen, digital rectal examination and transrectal ultrasonography: a meta-analysis for this diagnostic triad of prostate cancer in symptomatic korean men Yonsei Med J, 46(3): 414-24.
44. S. Sozen, S. Eskicorapci, B. Kupeli, L. Irkilata, M. Altinel, G. Ozer, C. Uygur, T. Alkibay and H. Ozen 2005. Complexed prostate specific

antigen density is better than the other PSA derivatives for detection of prostate cancer in men with total PSA between 2.5 and 20 ng/ml: results of a prospective multicenter study Eur Urol, 47(3): 302-7.
45. G. Stephan, M. Stroebel, A. Heinau, et al. The ratio of prostate-specific antigen (PSA) to prostate volume (PSA density) as a parameter to improve the detection of prostate carcinoma in PSA values in the range of < 4 ng/mL. Cancer. 2005 104(5): 993-1003.
46. A. J. Stephenson, M. W. Kattan, J. A. Eastham, F. J. Bianco, Jr., O. Yossepowitch, A. J. Vickers, E. A. Klein, D. P. Wood and P. T. Scardino 2009. Prostate cancer-specific mortality after radical prostatectomy for patients treated in the prostate-specific antigen era J Clin Oncol, 27(26): 4300-5.
47. S. A. Tomlins, S. M. Aubin, J. Siddiqui, R. J. Lonigro, L. Sefton-Miller, S. Miick, S. Williamsen, P. Hodge, J. Meinke, A. Blase, Y. Penabella, J. R. Day, R. Varambally, B. Han, D. Wood, L. Wang, M. G. Sanda, M. A. Rubin, D. R. Rhodes, B. Hollenbeck, K. Sakamoto, J. L. Silberstein, Y. Fradet, J. B. Amberson, S. Meyers, N. Palanisamy, H. Rittenhouse, J. T. Wei, J. Groskopf and A. M. Chinnaiyan 2011. Urine TMPRSS2:ERG Fusion Transcript Stratifies Prostate Cancer Risk in Men with Elevated Serum PSA Sci Transl Med, 3(94): 94ra72.
48. M. Verma, P. Patel. Biomarkers in prostate cancer epidemiology. Cancers. 2011 3(4):3773-98.
49. V. Vlaeminck-Guillem, A. Ruffion, J. Andre, M. Devonec and P. Paparel 2010. Urinary prostate cancer 3 test: toward the age of reason? Urology, 75(2): 447-53.
50. Y. Wang, G. Sun, J. G. Pan, Z. J. Guo and T. Li 2006. Performance of tPSA and f/tPSA for prostate cancer in Chinese. A systematic review and meta-analysis Prostate Cancer Prostatic Dis, 9(4): 374-8.
51. J. F. Ward 2006. Can PSA velocity serve as a surrogate endpoint in trials of hormone-refractory, metastatic prostate cancer? Nat Clin Pract Urol, 3(6): 310-1.
52. T. Yuasa, N. Tsuchiya, T. Kumazawa, T. Inoue, S. Narita, M. Saito, Y. Horikawa, S. Satoh and T. Habuchi 2008. Characterization of prostate cancer detected at repeat biopsy BMC Urology, 8(1).

Appendix E. Quality Assessment Criteria and Category Definitions for Nonrandomized Comparative Intervention Studies[84,85]

- Were the sample definition and selection prospective or retrospective?
- Were inclusion/exclusion criteria clearly described?
- Were participants selected to be representative?
- Was there an attempt to balance groups by design?
- Were baseline prognostic characteristics clearly described and groups shown to be comparable?
- Were interventions clearly specified?
- Were participants in treatment groups recruited within the same time period?
- Was there an attempt by investigators to allocate participants to treatment groups in an attempt to minimize bias?
- Were concurrent/concomitant treatments clearly specified and given equally to treatment groups?
- Were outcome measures clearly valid, reliable, and equally applied to treatment groups?
- Were outcome assessors blinded?
- Was the length of followup adequate?
- Was subject attrition below an overall high level (<20 percent)?
- Was the difference in attrition between groups below a high level (<15 percent)?
- Did the analysis of outcome data incorporate a method for handling confounders such as statistical adjustment?

The rating of intervention studies encompassed three quality categories:

- **Good** studies meet all criteria; comparable groups were assembled initially and maintained throughout the study (followup at least 80 percent); reliable and valid measurement instruments were used and applied equally to the groups; interventions are spelled out clearly; all important outcomes were considered; appropriate attention was given to confounders in analyzing data.
- **Fair** studies had any or all of the following problems, but without the fatal flaws noted in the "poor" category below; comparable groups were assembled initially, but some questions remain about whether some (although not major) differences occurred with followup; measurement instruments were acceptable (although not the best) and were generally applied equally; some, but not all, important outcomes were considered; some, but not all, potential confounders were accounted for.
- **Poor** studies have any of the following fatal flaws; groups assembled initially were not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments were used or not applied at all equally among groups (including not masking outcome assessment); key confounders were given little or no attention.

Appendix F. Quality Assessment of Studies Addressing KQ 1/KQ 2 and KQ 3

Table F-1. PCA3 testing for the diagnosis and management of prostate cancer—quality of studies addressing KQs 1 and 2

QUADAS Criteria	Adam 2011[1]	Ankerst 2010[2]	Aubin 2010[3]	Auprich 2011[4]	Bollito 2012[5]
Representative study subjects	Yes	Yes	Yes	Yes	Uncertain[a,b]
Selection criteria clear	Yes	Yes	Yes	Yes	Yes
Reference standard correctly identifies PCa	Yes	Yes	Yes	Yes	Yes
Acceptable period between index test and reference standard	Yes	Yes	Yes	Yes	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes	Yes	Yes
Index test adequately described	Yes	No	Yes	Yes	Yes
Reference standard adequately described	Yes	Yes	Yes	Yes	Yes
Index test results interpreted without knowledge of reference standard results[c]	Yes	Uncertain	Yes	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[c]	Yes	Uncertain	Uncertain	Uncertain	Uncertain
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Yes	Yes	Yes
Uninterpretable/intermediate test results were reported	No	No	No	No	Uncertain
Withdrawals from the study were explained	Yes	None reported	None reported	None reported	Yes
Added criteria:					
Reference standard results interpreted without knowledge of the tPSA test	No	No	No	No	No
Partial verification bias[d]	Yes	Yes	Yes[e]	Yes	Yes
Score	**Poor**	**Poor**	**Poor**	**Poor**	**Poor**

The *index test* is PCA3 except for the last three lines, which refer to the tPSA comparator. The *reference standard i*s biopsy.

Table F-1. PCA3 testing for the diagnosis and management of prostate cancer – quality of studies addressing KQs 1 and 2 (continued)

QUADAS Criteria	Cao 2011[6]	de la Taille 2011[7]	Deras 2008[8]	FDA Summary[9]	Ferro 2012[10]
Representative study subjects	Uncertain[a]	Yes	Yes	No	Uncertain[b]
Selection criteria clear	No	Yes	Yes	No	Yes
Reference standard correctly identifies PCa	Yes	Yes	Yes	Yes	Yes
Acceptable period between index test and reference standard	Yes	Yes	Yes	Yes	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes	Yes	Yes
Index test adequately described	Yes	Yes	Yes	Yes	Yes
Reference standard adequately described	Yes	Yes	Yes	Yes	Yes
Index test results interpreted without knowledge of reference standard results[c]	Uncertain	Yes	Uncertain	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[c]	Uncertain	Uncertain	Yes	Uncertain	Yes
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Yes	Yes	Yes
Uninterpretable/intermediate test results were reported	No	No	No	No	No
Withdrawals from the study were explained	Yes	Uncertain for %fPSA	Yes	Yes	None reported
Added criteria:					
Reference standard results interpreted without knowledge of the tPSA test	No	No	No	Uncertain	No
Partial verification bias[d]	Yes	Yes[e]	Yes	Uncertain but likely	Yes[e]
Score	**Poor**	**Poor**	**Poor**	**Poor**	**Poor**

The *index test* is PCA3 except for the last three lines, which refer to the tPSA comparator. The *reference standard i*s biopsy.

Table F-1. PCA3 testing for the diagnosis and management of prostate cancer—quality of studies addressing KQs 1 and 2 (continued)

QUADAS Criteria	Goode 2012[11]	Hessels 2010[12]	Mearini 2010[13]	Nyberg 2010[14]	Ochai 2011[15]
Representative study subjects	Yes	Yes	Uncertain[a]	Yes	Yes
Selection criteria clear	Yes	Yes	Yes	Yes	Yes
Reference standard correctly identifies PCa	Yes	Yes	Yes	Yes	Yes
Acceptable period between index test and reference standard	Yes	Yes	Yes	Uncertain	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes	Yes	Yes
Index test adequately described	Yes	Yes	No	Yes	Yes
Reference standard adequately described	Yes	Yes	No	Uncertain	Yes
Index test results interpreted without knowledge of reference standard results[c]	Yes	Uncertain	Uncertain	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[c]	Yes	Uncertain	Uncertain	Uncertain	Uncertain
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Uncertain	Yes	Yes
Uninterpretable/intermediate test results were reported	No	No	Yes	No	No
Withdrawals from the study were explained	None reported	Yes	None reported	None reported	Yes
Added criteria:					
Reference standard results interpreted without knowledge of the tPSA test	Uncertain	No	No	No	No
Partial verification bias[d]	Yes	Yes	Yes	Yes	Yes
Score	**Poor**	**Poor**	**Poor**	**Poor**	**Poor**

The *index test* is PCA3 except for the last three lines, which refer to the tPSA comparator. The *reference standard i*s biopsy.

Table F-1. PCA3 testing for the diagnosis and management of prostate cancer—quality of studies addressing KQs 1 and 2 (continued)

QUADAS Criteria	Ouyang 2009[16]	Pepe 2012[17]	Perdona 2011[18]	Ploussard 2010[19]	Rigau 2008[20]
Representative study subjects	Uncertain[a]	Uncertain	Yes	Yes	Yes
Selection criteria clear	No	No	Yes	Yes	Yes
Reference standard correctly identifies PCa	Yes	Yes	Yes	Yes	Yes
Acceptable period between index test and reference standard	Yes	Yes	Yes	Yes	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes	Yes	Yes
Index test adequately described	Yes	Yes	Yes	Yes	Yes
Reference standard adequately described	No	Yes	Yes	Yes	Yes
Index test results interpreted without knowledge of reference standard results[c]	Uncertain	Uncertain	Uncertain	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[c]	Uncertain	Uncertain	Uncertain	Uncertain	Uncertain
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Yes	Yes	Yes
Uninterpretable/intermediate test results were reported	No	No	No	No	No
Withdrawals from the study were explained	Yes	None reported	None reported	Yes	Yes
Added criteria:					
Reference standard results interpreted without knowledge of the tPSA test	No	Uncertain	No	No	No
Partial verification bias[d]	Yes	Yes	Yes[e]	Yes[e]	Yes
Score	**Poor**	**Poor**	**Poor**	**Poor**	**Poor**

The *index test* is PCA3 except for the last three lines, which refer to the tPSA comparator. The *reference standard i*s biopsy.

Table F-1. PCA3 testing for the diagnosis and management of prostate cancer—quality of studies addressing KQs 1 and 2 (continued)

QUADAS Criteria	Roobol 2010[21]	Schilling 2011[22]	Wang 2009[23]	Wu 2012[24]
Representative study subjects	Yes	No	Yes	Yes
Selection criteria clear	Yes	Yes	Yes	Uncertain
Reference standard correctly identifies PCa	Yes	Yes	Yes	Yes
Acceptable period between index test and reference standard	Yes	Uncertain	Uncertain	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes	Yes
Index test adequately described	Yes	Yes	Yes	Yes
Reference standard adequately described	Yes	Yes	Yes	Yes
Index test results interpreted without knowledge of reference standard results[c]	Uncertain	Uncertain	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[c]	Uncertain	Uncertain	Uncertain	Uncertain
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Yes	Yes
Uninterpretable/intermediate test results were reported	No	No	No	No
Withdrawals from the study were explained	Yes	Yes	Yes	None reported
Added criteria:				
Reference standard results interpreted without knowledge of the tPSA test	No	No	No	N
Partial verification bias[d]	Yes	Yes	Yes	Yes
Score	**Poor**	**Poor**	**Poor**	**Poor**

The *index test* is PCA3 except for the last three lines, which refer to the tPSA comparator. The *reference standard i*s biopsy.

[a] 'Uncertain' indicates that information was not provided to determine the answer to the question; it does not mean that the answer to the question is 'No'.
[b] Uncertain because these studies excluded men with positive DRE.
[c] Information on blinding is often not reported, so the category of Uncertain does not mean that blinding did not occur.
[d] Partial verification bias because men with elevated tPSA and/or positive digital rectal exam who did not accept biopsy are considered missing.
[e] There is an additional partial verification bias in this study due to a limitation on the range of tPSA values that are included (i.e., the so called 'grey zone' results in the range of 2.5 to 10 or 20 ng/mL).

Table F-2. PCA3 testing for the diagnosis and management of prostate cancer—quality of studies addressing KQ3

QUADAS Criteria	Auprich 2011[25]	Durand 2012[26]	Kasuda 2011[27]	Liss 2011[28]
Representative study subjects	Yes	Yes	Uncertain[a]	Yes
Selection criteria clear	Yes	Yes	Yes	Yes
Reference standard correctly identifies PCa or other outcome	Yes	Yes	Uncertain	Yes
Acceptable period between index test and reference standard	Yes	Yes	Yes	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes	Yes
Index test adequately described	Yes	Yes	Uncertain	Yes
Reference standard adequately described	Yes	Yes	No	Yes
Index test results interpreted without knowledge of reference standard results[b]	Uncertain	Uncertain	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[b]	Uncertain	Uncertain	Uncertain	Uncertain
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Yes	Yes
Uninterpretable/intermediate test results were reported	No	No	Yes	No
Withdrawals from the study were explained	Yes	None reported	Uncertain	Yes
Added criterion: *Clinical followup*	No	No	Yes	No
Score	**Poor**	**Poor**	**Poor**	**Poor**

Table F-2. PCA3 testing for the diagnosis and management of prostate cancer—quality of studies addressing KQ3 (continued)

QUADAS Criteria	Nakanishi 2008[29]	Ploussard 2010[30]	Tosian 2010[31]
Representative study subjects	Yes	Yes	Yes
Selection criteria clear	Yes	Yes	Yes
Reference standard correctly identifies PCa or other outcome	Yes	Yes	Uncertain
Acceptable period between index test and reference standard	Yes	Yes	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes
Index test adequately described	Yes	Yes	Yes
Reference standard adequately described	Yes	Yes	No
Index test results interpreted without knowledge of reference standard results[b]	Uncertain	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[b]	Uncertain	Uncertain	Uncertain
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Uncertain
Uninterpretable/intermediate test results were reported	No	No	Yes
Withdrawals from the study were explained	None reported	None reported	No
Added criterion: *Clinical followup*	No	No	Yes
Score	**Poor**	**Poor**	**Poor**

Table F-2. PCA3 testing for the diagnosis and management of prostate cancer—quality of studies addressing KQ3 (continued)

QUADAS Criteria	van Poppel 2011[32]	Vlaeminck-Guillem 2011[33]	Whitman 2008[34]
Representative study subjects	Yes	Yes	Yes
Selection criteria clear	Yes	Yes	Yes
Reference standard correctly identifies PCa or other outcome	Yes	Yes	Yes
Acceptable period between index test and reference standard	Yes	Yes	Yes
All or described subset diagnosed by reference standard	Yes	Yes	Yes
All patients received reference standard, regardless of index result	Yes	Yes	Yes
Reference standard was independent of the index test	Yes	Yes	Yes
Index test adequately described	Yes	Yes	Yes
Reference standard adequately described	Yes	Yes	Yes
Index test results interpreted without knowledge of reference standard results[b]	Yes	Uncertain	Uncertain
Reference standard results interpreted without knowledge of index test[b]	Uncertain	Yes	Uncertain
Clinical data for interpretation of test results were the same as those expected in practice	Yes	Yes	Yes
Uninterpretable/intermediate test results were reported	No	No	No
Withdrawals from the study were explained	None reported	Uncertain	None reported
Added criterion: *Clinical followup*	No	No	No
Score	**Poor**	**Poor**	**Poor**

[a] 'Uncertain' indicates that information was not provided to determine the answer to the question; it does not mean that the answer to the question is 'No'.
[b] Information on blinding is often not reported, so the category of 'Uncertain' does not mean that blinding did not occur.

References

1. Adam A, Engelbrecht MJ, Bornman MS, et al. The role of the PCA3 assay in predicting prostate biopsy outcome in a South African setting. BJU Intl. 2011 Apr 20; PMID: 21507188.
2. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008 Oct;180(4):1303-8; discussion 08. PMID: 18707724.
3. Aubin SM, Reid J, Sarno MJ, et al. PCA3 molecular urine test for predicting repeat prostate biopsy outcome in populations at risk: validation in the placebo arm of the dutasteride REDUCE trial. J Urol. 2010 Nov;184(5):1947-52. PMID: 20850153.
4. Auprich M, Augustin H, Budaus L, et al. A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. BJU Int. 2012 Jun;109(11):1627-35. PMID: 21939492.
5. Bollito E, De Luca S, Cicilano M, et al. Prostate cancer gene 3 urine assay cutoff in diagnosis of prostate cancer: a validation study on an Italian patient population undergoing first and repeat biopsy. Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology. 2012 Apr;34(2):96-104. PMID: 22611765.
6. Cao DL, Ye DW, Zhang HL, et al. A multiplex model of combining gene-based, protein-based, and metabolite-based with positive and negative markers in urine for the early diagnosis of prostate cancer. Prostate. 2011 May 15;71(7):700-10. PMID: 20957673.
7. de la Taille A, Irani J, Graefen M, et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. J Urol. 2011 Jun;185(6):2119-25. PMID: 21496856.
8. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol. 2008 Apr;179(4):1587-92. PMID: 18295257.
9. U.S. Food and Drug Administration. Summary of Safety and Effectiveness Data: PROGENSA PCA3 Assay, 2012. Accessed July 16, 2012. http://www.accessdata.fda.gov/cdrh_docs/pdf10/P100033b.pdf.
10. Ferro M, Bruzzese D, Perdona S, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta. 2012 Aug 16;413(15-16):1274-8. PMID: 22542564.
11. Goode RR, Marshall SJ, Duff M, et al. Use of PCA3 in detecting prostate cancer in initial and repeat prostate biopsy patients. Prostate. 2012 May 14; PMID: 22585386.
12. Hessels D, van Gils MP, van Hooij O, et al. Predictive value of PCA3 in urinary sediments in determining clinico-pathological characteristics of prostate cancer. Prostate. 2010 Jan 1;70(1):10-6. PMID: 19708043.
13. Mearini E, Antognelli C, Del Buono C, et al. The combination of urine DD3(PCA3) mRNA and PSA mRNA as molecular markers of prostate cancer. Biomarkers. 2009 Jun;14(4):235-43. PMID: 19489685.
14. Nyberg M, Ulmert D, Lindgren A, et al. PCA3 as a diagnostic marker for prostate cancer: a validation study on a Swedish patient population. Scand J Urol Nephrol. 2010 Dec;44(6):378-83. PMID: 20961267.
15. Ochiai A, Okihara K, Kamoi K, et al. Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. Int J Urol. 2011 Mar;18(3):200-5. PMID: 21332814.
16. Ouyang B, Bracken B, Burke B, et al. A duplex quantitative polymerase chain reaction assay based on quantification of alpha-methylacyl-CoA racemase transcripts and prostate cancer antigen 3 in urine sediments improved diagnostic accuracy for prostate cancer. J Urol. 2009 Jun;181(6):2508-13; discussion 13-4. PMID: 19371911.
17. Pepe PAragona F. PCA3 score vs PSA free/total accuracy in prostate cancer diagnosis at repeat saturation biopsy. Anticancer Res. 2011 Dec;31(12):4445-9. PMID: 22199313.
18. Perdona S, Cavadas V, Di Lorenzo G, et al. Prostate cancer detection in the "grey area" of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. European urology. 2011 Jan;59(1):81-7. PMID: 20947244.
19. Ploussard G, Haese A, Van Poppel H, et al. The prostate cancer gene 3 (PCA3) urine test in men with previous negative biopsies: does free-to-total prostate-specific antigen ratio influence the performance of the PCA3 score in predicting positive biopsies? BJU Int. 2010 Oct;106(8):1143-7. PMID: 20230386.

20. Rigau M, Morote J, Mir MC, et al. PSGR and PCA3 as biomarkers for the detection of prostate cancer in urine. Prostate. 2010 Dec 1;70(16):1760-7. PMID: 20672322.
21. Roobol MJ, Schroder FH, van Leeuwen P, et al. Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: exploring the value of PCA3 for a first-line diagnostic test. European urology. 2010 Oct;58(6):475-81. PMID: 20637539.
22. Schilling D, Hennenlotter J, Munz M, et al. Interpretation of the prostate cancer gene 3 in reference to the individual clinical background: implications for daily practice. Urol Int. 2010 85(2):159-65. PMID: 20424427.
23. Wang R, Chinnaiyan AM, Dunn RL, et al. Rational approach to implementation of prostate cancer antigen 3 into clinical care. Cancer. 2009 Sep 1;115(17):3879-86. PMID: 19517474.
24. Wu AK, Reese AC, Cooperberg MR, et al. Utility of PCA3 in patients undergoing repeat biopsy for prostate cancer. Prostate Cancer Prostatic Dis. 2012 Mar;15(1):100-5. PMID: 22042252.
25. Auprich M, Chun FK, Ward JF, et al. Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. Euro Urol. 2011 Jan;59(1):96-105. PMID: 20980098.
26. Durand X, Xylinas E, Radulescu C, et al. The value of urinary prostate cancer gene 3 (PCA3) scores in predicting pathological features at radical prostatectomy. BJU Int. 2012 Jul;110(1):43-49. PMID: 22221521.
27. Kusuda Y, Miyake H, Kurahashi T, et al. Assessment of optimal target genes for detecting micrometastases in pelvic lymph nodes in patients with prostate cancer undergoing radical prostatectomy by real-time reverse transcriptase-polymerase chain reaction. Urol Oncol. 2011 May 18; PMID: 21600799.
28. Liss MA, Santos R, Osann K, et al. PCA3 molecular urine assay for prostate cancer: association with pathologic features and impact of collection protocols. World J Urol. 2011 Oct;29(5):683-8. PMID: 21152924.
29. Nakanishi H, Groskopf J, Fritsche HA, et al. PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. J Urol. 2008 May;179(5):1804-9; discussion 09-10. PMID: 18353398.
30. Ploussard G, Durand X, Xylinas E, et al. Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance. European urology. 2011 Mar;59(3):422-9. PMID: 21156337.
31. Tosoian JJ, Loeb S, Kettermann A, et al. Accuracy of PCA3 measurement in predicting short-term biopsy progression in an active surveillance program. J Urol. 2010 Feb;183(2):534-8. PMID: 20006883.
32. van Poppel H, Haese A, Graefen M, et al. The relationship between Prostate CAncer gene 3 (PCA3) and prostate cancer significance. BJU Int. 2012 Feb;109(3):360-6. PMID: 21883822.
33. Vlaeminck-Guillem V, Devonec M, Colombel M, et al. Urinary PCA3 score predicts prostate cancer multifocality. J Urol. 2011 Apr;185(4):1234-9. PMID: 21334023.
34. Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. J Urol. 2008 Nov;180(5):1975-8; discussion 78-9. PMID: 18801539.

Appendix G. DistillerSR Abstract and Title Screening Form

Is the article in English, or was an English translation available from the journal?

☐ Yes

☐ No

☐ Uncertain

Were the study participants men in any of these specific populations:

- men at risk for prostate cancer based on elevated PSA (lowest cutoff > 2.5 ng/mL) and/or positive DRE?
- men at risk for prostate cancer based on elevated PSA and/or positive DRE, and one or more negative prostate biopsies?
- men with a positive prostate biopsy?

☐ Yes

☐ No

☐ Uncertain

Does at least one of the study descriptions below apply?

- A matched study or systematic review/meta-analysis of PCA3 and one or more designated comparators(s) (tPSA, %fPSA, cPSA, PSA velocity, validated nomograms) used to inform a decision about initial or repeat prostate biopsy.
- A matched study or systematic review/meta-analysis of PCA3 and comparators (Gleason score, tumor stage/volume) used to assess prognosis/tumor aggressiveness.
- A matched study or systematic review/meta-analysis of PCA3 alone or in combination with comparators (tPSA, PSA velocity) as part of triage for active surveillance or aggressive treatment.
- A systematic review and/or meta-analysis of matched studies of PSA and one or more comparator(s) used to inform decisions about initial or repeat prostate biopsy.
- A systematic review, meta-analysis or matched study of any two or more comparators other than PSA used to inform decisions about initial or repeat prostate biopsy.

Note: Acceptable study designs included randomized or non-randomized controlled trials, prospective or retrospective cohort studies, diagnostic accuracy and case-control studies.

☐ Yes

☐ No

☐ Uncertain

Is this an unmatched diagnostic accuracy study (or systematic review/meta-analysis of such studies) on PCA3 or other comparators?

Select an Answer *[If Yes, paper was filed for future reference but ineligible for abstraction.]*

Appendix H. DistillerSR Full-Text Article Screening Form

Question Text	Type	Question Header	Answer Text	Answer Headers
Was there an English language article or a translation available from the journal?	Radio	English language	Yes, No, Uncertain	Yes, No, Uncertain
Were the study participants men in any of these specific populations: • men at risk for prostate cancer based on elevated PSA (lowest cutoff > 2.5 ng/mL) and/or positive DRE? • men at risk for prostate cancer based on elevated PSA and/or positive DRE, and one or more negative biopsies? • men with a positive prostate biopsy?	Radio	Appropriate study participants	Yes, No, Uncertain	Yes, No, Uncertain
Does at least one of the study descriptions below apply? • A matched study or systematic review/meta-analysis of PCA3 and one or more designated comparators(s) (tPSA, %fPSA, cPSA, PSA velocity, validated nomograms) used to inform a decision about initial or repeat prostate biopsy. • A matched study or systematic review/meta-analysis of PCA3 and comparators (Gleason score, tumor stage/volume) used to assess prognosis/tumor aggressiveness. • A matched study or systematic review/meta-analysis of PCA3 alone or in combination with comparators (tPSA, PSA velocity) as part of triage for active surveillance or aggressive treatment. • A systematic review and/or meta-analysis of matched studies of PSA and one or more comparator(s) used to inform decisions about initial or repeat prostate biopsy. • A systematic review, meta-analysis or matched study of any two or more comparators other than PSA used to inform decisions about initial or repeat prostate biopsy. • A systematic review, meta-analysis or comparative study evaluating long-term outcomes. Note: Acceptable study designs included randomized or non-randomized controlled trials, prospective or retrospective cohort studies, diagnostic accuracy and case-control studies.	Radio	Appropriate study description	Yes, No, Uncertain	Yes, No, Uncertain
Did study participants undergo prostate biopsy or prostatectomy as part of a diagnostic workup for prostate cancer?	Radio	Biopsy or RP	Yes, No, Uncertain	Yes, No, Uncertain
Did the biopsy include a minimum of 6 cores?	Radio	≥ 6 biopsy cores	Yes, No	Yes, No

			Uncertain	Uncertain
Did the study report provide one or more of the criteria used to characterize tumors (e.g., Gleason score, tumor burden, staging)? *[Required only for KQ3]*	Radio	Tumor characteristics	Yes, No, Not Applicable, Uncertain	Yes, No, Not Applicable, Uncertain
For patients who were candidates for initial or repeat biopsy, did the study provide: • estimate(s) of diagnostic accuracy for prostate cancer; • short-term outcomes (e.g., decision to biopsy or not, medical and psychological complications of biopsy); and/or • long-term outcomes (e.g., mortality, morbidity, quality of life, harms) based on intervention. OR For patients who were biopsy positive and being triaged for active surveillance or aggressive treatment, did the study provide: • estimates of diagnostic accuracy for aggressive disease; • short-term outcomes (e.g., decisions regarding aggressive surveillance or treatment); and/or • long-term outcomes (e.g., mortality, morbidity, quality of life, harms) based on intervention.	Radio	Outcomes reported	Yes, No, Uncertain	Yes, No, Uncertain

Appendix I. DistillerSR Data Extraction Forms

Study Description

Question Text	Answers Text
First Author (Last, First Initial):	
Year of Publication:	
Study design:	RCT, Non-randomized controlled trial, Non-randomized comparative study, Systematic review/MA, Diagnostic accuracy study, Case-control study
Is this a matched study?	Yes, No
Does this study report systematic review data within it?	Yes, No
Which key question(s) does this article address? KQ1:In patients with elevated PSA and/or an abnormal DRE who are candidates for initial prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests (e.g., elevated total PSA values, decreased percent-percent free PSA levels, elevated PSA velocities, complexed PSA, or externally validated nomograms) with regard to diagnostic accuracy, intermediate outcomes and/or long-term health outcomes? KQ2: In patients with elevated PSA and/or an abnormal DRE who are candidates for repeat prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard screening tests (e.g., elevated total PSA values, decreased percent percent free PSA levels, elevated PSA velocities, complexed PSA, or externally validated nomograms) with regard to diagnostic accuracy, intermediate outcomes and/or long-term health outcomes? KQ3: In patients with a positive biopsy for prostate cancer who are being evaluated to distinguish between indolent and aggressive disease, what is the effectiveness of using PCA3 testing alone, or in combination with the standard prognostic workup (e.g., tumor volume, Gleason score, clinical staging) or monitoring tests (e.g., PSA, PSA velocity) with regard to diagnostic accuracy, intermediate outcomes and/or long-term health outcomes?	KQ1, KQ2, KQ3, All, None
Does this article report data on men who are first time biopsies, repeat biopsies, mixed or uncertain?	First biopsies, Repeat biopsies, Mixed, Uncertain, Not applicable
Title:	
Is there more than one study site?	Yes, No
Total number of study sites	
List the name of each study site below:	Institution 1-7
Name of Study Site:	
Start Date and End Date (mm/dd/yy):	
Number at enrollment:	
Setting(s):	Hospital/clinic, Referrral center, Screening FU, Other
Follow-up assessed?	Yes, No
Follow-up Number:	

Follow-up Percent (%):	
BOTH not provided?	Yes
Follow-up losses explained?	Yes, No-Not Provided
Follow-up (months):	Mean, SD, Median, Range, IQR
Source(s) of Funding:	Not Provided, Not Externally Funded, Departmental, Industry, Government, Private Funding, Other
Author-industry relationship disclosures:	None, Yes – describe, Not provided
Does this article reference other publications where this dataset, subsets of this dataset, or this study population were used?	Yes – Indicate reference, No
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	
FACT-CHECKING FOR THIS ARTICLE HAS BEEN COMPLETED BY:	
Final fact-checking completed on (mm/dd/yy):	
THIS ARTICLE IS NOT A PCA3 ARTICLE?	Yes, No

Participant Characteristics & Inclusion Criteria

Question Text	Answers Text
Age:	Mean, SD, Median, Range, IQR
Men on meds excluded:	Yes, No, Not specified
Race: African-Americans N(%):	
Caucasian N(%):	
Hispanic N(%):	
Asian N(%):	
Other N(%):	
Are there men with benign prostatic hyperplasia (BPH)?: Are there men with high-grade prostate intraepithelial neoplasia (HGPIN)?: Are there men with atypical small acinar proliferation (ASAP)?: Are there men with prostatitis?:	Yes –N, %, No-Excluded, Unknown
Were results of DRE reported outside of inclusion?	Yes, No
Number of men who had DRE Test:	
Number and % DRE positive:	
Number and % DRE negative:	
PSA testing in group?	Yes, No
tPSA cutoff 1-4: Number tested, Number positive, % positive	
Were men included based on their DRE test?	Yes-Number, No, Unknown
Number and % DRE positive	
One or more previous negative biopsies:	Yes-N, %, No

Positive Family History:	Yes-N, %, No, Unknown
African-American:	Yes-N, %, No, Unknown
Age Cutoff:	Yes-Specify, N, %, No, Unknown
Scheduled for biopsy—specific reason for biopsy unknown:	Yes-N, %, No, Unknown
ABSTRACTOR COMMENTS If you would like to leave a comment pertaining to the information above indicate your name below:	

Specimen/Assay Description

Question Text	Answers Text
Collection by attentive prostate massage:	Yes, No, Not provided
Specimen:	Urine – sedimented, Urine – unsedimented, Prostatic ejaculate
Name of transport media:	
Transport time (hours):	
Holding temperature (Celsius):	
Storage temperature (-C):	
Method used:	Quantitative Real Time PCR, Nucleic acid sequence based amplification, Transcription-mediated Amplication
Assay used identified as:	1^{st}, 2^{nd} or 3^{rd} generation test
	uPM3, PCA3Plus, Progensa, Aptima, “Gen-Probe test”, Test not specified
Housekeeping gene:	PSA mRNA, Other, Not provided
Result unit:	PCA3 Score, Other
Testing blinded to outcomes:	Yes, No, Not provided
Number and percent of samples with insufficient mRNA:	
Number and percent of reportable results:	
PCA3 losses explained?	Yes, No
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

PCA3 by Cutoff

Question Text	Answer Text
This page does not pertain to this study?	Yes (move to next form), No (complete #2)
This study has more than one PCA3 cutoff?	Yes (complete #3 and #4), No (skip #3)
Indicate PCA3 cutoff number:	PCA2 cutoff 1-6
For each, PCA3 cutoff value:	

Negative biopsy summary for PCA3::	Mean, SD, Median, Range, IQR
Negative biopsy summary for PCA3:	Mean, SD, Median, Range, IQR
Number Tested:	
Raw data:	TP = Test positive, biopsy positive, FP = Test positive, biopsy negative, FN = Test negative, biopsy positive, TN = Test negative, biopsy negative
Performance data	Sensitivity, CI, p-value; Specificity, CI, p-value; Positive predictive value, CI, p-value, Negative predictive value, CI, p-value, Odds ratio, Likelihood ratio
Was a univariate analysis conducted for Diagnostic Odds Ratio (DOR)?	Yes-Predictive Accuracy, p-value, No
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

Total PSA by Cutoff (Identical format used for total PSA, percent free PSA, PSA density, PSA doubling time, PSA velocity, complexed PSA, externally validated nomograms)

Question Text	Answer Text
This page does not pertain to this study?	Yes (move to next form), No (complete #2)
This study has more than one tPSA Cutoff?	Yes (complete #3 and #4), No (skip #3)
Indicate tPSA Cutoff number:	tPSA Cutoff 1-6
For each: tPSA Cutoff Value:	
Negative biopsy summary for PSA::	Mean, SD, Median, Range, IQR
Negative biopsy summary for PSA:	Mean, SD, Median, Range, IQR
Number Tested:	
Raw data:	TP = Test positive, biopsy positive, FP = Test positive, biopsy negative, FN = Test negative, biopsy positive, TN = Test negative, biopsy negative
Performance data	Sensitivity, CI, p-value; Specificity, CI, p-value; Positive predictive value, CI, p-value, Negative predictive value, CI, p-value, Odds ratio, Likelihood ratio
If compared to PCA3 identify method for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
p-value:	
Was a univariate analysis conducted for Diagnostic Odds Ratio (DOR)?	Yes-Predictive Accuracy, p-value, No
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

Prostate biopsy and prostatectomy

Question Text	Answer Text
Indicate for which the total number is given:	Biopsies, Prostatectomies, Both
Indicate total N reportable biopsies:	
Indicate total N prostatectomies:	
Biopsy losses explained?	Yes, No, Unknown
Number of cores per biopsy:	
Cores per biopsy:	Mean, SD, Median, Range, IQR
Interpretation blinded to study categories:	Yes, No, Not provided
Positive biopsies:	N, %
Percent cores with cancer:	% or Not Provided
Cores with cancer > 33%:	Number, %
% cancer per core:	Mean %, SD, Median, Range, IQR
Gleason score: 6; 7; >6; >7; ≥ 7:	N, %
Gleason Score cutoff not listed above:	Yes, No
Other Gleason score cutoffs: 1-3	Name cutoff, N, %
PSA Density	Mean, SD, Median, Range, IQR
Biopsy Pathological stages: T1c; T2a; T2b; T2c; T3	N, %
Other biopsy pathological stage:	Name stage, N, %
Prostate Volume (cc):	Mean, SD, Median, Range, IQR
Are there “Insignificant findings” based on Epstein criteria (clinical stage T1c, PSA density <0.15, Gleason score <6, presence of PCa in fewer than 3 cores, <33% of cores are positive):	Meets Epstein Criteria- N, %; Does NOT Meet Epstein Criteria
Radical prostatectomy staging: pT0; pT1, pT2; pT3	N, %
Extracapsular extension:	N, %
Perineural invasion:	N, %
Nodal involvement:	N, %
Seminal vesicle invasion:	N, %
Abnormal MRI:	N, %
Abnormal TRUS:	N, %
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

Gleason Score

Question Text	Answer Text
This page does not pertain to this study?	Yes (move to next form), No (complete #2)
This study has more than one Gleason score Cutoff?	Yes (complete #3 and #4), No (skip #3)
Indicate Gleason Score Cutoff number:	Cutoff 1-6
Indicate Gleason Score Cutoff Value:	

Number of men WITH this biopsy result who are having a PCA3 test:	
PCA3 value Mean:	
If compared to PCA3, identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
p-value:	
PCA3 value Standard Deviation (SD):	
PCA3 value Median:	
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
p-value:	
PCA3 value Range:	
PCA3 value IQR:	
Area Under the Curve (AUC):	Value, CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
Was a univariate analysis conducted for Diagnostic Odds Ratio (DOR)?	Yes-Predictive Accuracy, p-value, No
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

Pathological Staging

Question Text	Answer Text
This page does not pertain to this study?	Yes (move to next form), No (complete #2)
This study presents information on more than one pathological staging result:	Yes (complete #3 and #4 and skip to Table J1.), No (skip #3 and complete #4)
Identify path result number:	1-6
Identify pathological staging result:	T1c, T2, T2a, T2b, T2c, Other
Number of men WITH this biopsy result who are having a PCA3 test:	
PCA3 value Mean:	
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
p-value:	
PCA3 value Standard Deviation (SD):	
PCA3 value Median:	
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other

p-value:	
PCA3 value Range:	
PCA3 value IQR:	
Area Under the Curve (AUC):	Value, CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

Epstein Criteria

Question Text	Answer Text
This page does not pertain to this study?	Yes (move to next form), No (complete #2)
Does this study more than one set of criteria?	Yes (complete #3 and #4), No (skip #3)
Describe components of Epstein criteria 1-5	
Indicate Epstein Criteria category (category defined by each set of criteria):	Indolent, Significant, Other
Is the Epstein criteria modified?	Yes, No
Criteria Set 1-5:	Describe criteria
Number of men WITH this biopsy result having a PCA3 test:	
PCA3 value Mean:	
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
p-value:	
PCA value Standard Deviation (SD):	
PCA3 value Median:	
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
p-value:	
PCA3 value Range:	
PCA3 value IQR:	
Area Under the Curve (AUC):	Value, CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
p-value:	
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

AUC

Question Text	Answer Text
Test 1:	PCA3
Units:	
Reference standard:	Prostate biopsy, Prostatectomy
AUC Group 1:	Name
G 1. Area Under the Curve (AUC):	Value, 95% CI, p-value
Was AUC reported for more than one group?	Yes, No
AUC Group 2:	Name
G 2. Area Under the Curve (AUC):	Value, 95% CI, p-value
AUC Group 3:	Name
G 3. Area Under the Curve (AUC):	Value, 95% CI, p-value
AUC Group 4:	Name
G 4. Area Under the Curve (AUC):	Value, 95% CI, p-value
AUC Group 5:	Name
G 5. Area Under the Curve (AUC):	Value, 95% CI, p-value
Test 2-6	Names 2-6
Units:	
Reference standard:	Prostate biopsy, Prostatectomy
AUC Group 1:	Name
G 1. Area Under the Curve (AUC):	Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
Was AUC reported for more than one group?	Yes, No
AUC Group 2:	Name
G 2. Area Under the Curve (AUC):	Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
AUC Group 3:	Name
G 3. Area Under the Curve (AUC):	Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
AUC Group 4:	Name
G 4. Area Under the Curve (AUC):	Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
AUC Group 5:	Name

G 5. Area Under the Curve (AUC):	Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
Nomogram 1-3:	Name
Variables used in nomogram (must be 3 or more):	Age, family history, race, DRE result, previous negative biopsy, PCA3, tPSA, %fPSA, cPSA, PSA velocity, Gleason score, other
Method for score calculation:	
Reference Standard:	Prostate biopsy, Prostatectomy
Nomogram AUC Group 1:	Name, Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
Was AUC reported for more than one group?	Yes, No
Nomogram AUC Group 2	Name, Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
Nomogram AUC Group 3:	Name, Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
Nomogram AUC Group 4:	Name, Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
Nomogram AUC Group 5:	Name, Value, 95% CI, p-value
If compared to PCA3 identify test for comparison:	McNemar, Fisher's exact, paired t-test, Chi-square, Wilcoxon rank sum, Kruskal-Wallis, Pearson correlation, non-parametric comparison of AUC (DeLong), Other
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	
ENTER YOUR COMMENTS:	

Matched Analysis Data
Matched Study Data—Multivariate Analysis

Question Text	Answer Text
Is this study a matched analysis?	Yes (complete Section 1), No (skip to Section 2)
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

Section 1. Matched Analysis Data

Positive Biopsy:

	Comparator Positive	Comparator Negative
PCA3 Positive		
PCA3 Negative		

Negative Biopsy:

	Comparator Positive	Comparator Negative
PCA3 Positive		
PCA3 Negative		

Identify Comparator:

Section 2. Multivariate Logistical Regression Analysis

	Base ModelOdds Ratio (OR)	Base Model(95% CI)	Base Model (p-value)	Base Model + Group 1 (OR)	Base Model + Group 1 (95% CI)	Base Model + Group 1 (p-value)	Base Model + Group 2 (OR)	Base Model + Group 2 (95% CI)	Base Model + Group 2 (p-value)	Base Model + Group 3 (OR)	Base Model + Group 3 (95% CI)	Base Model + Group 3 (p-value)	Base Model + Group 4 (OR)	Base Model + Group 4 (95% CI)	Base Model + Group 4 (p-value)
Variable 1:															
Variable 2:															

Variable 3:															
Variable 4:															
Variable 5:															
Variable 6:															
Variable 7:															
Variable 8:															
Variable 9:															
Variable 10:															
Predictive Accuracy:															
Result OTHER:															

Identify Group 1 in the Model:

Identify Group 2 in the Model:

Identify Group 3 in the Model:

Identify Group 4 in the Model:

Quality - QUADAS

Question Text	Answers Text
Was the spectrum of patients representative of the patients who will receive the test in practice?	Yes, No, Uncertain
Were selection criteria clearly described?	Yes, No, Uncertain
Is the reference standard likely to classify the target condition correctly?	Yes, No, Uncertain
Is the period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Yes, No, Uncertain
Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?	Yes, No, Uncertain
Did patients receive the same reference standard regardless of the index test result?	Yes, No, Uncertain
Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard?)	Yes, No, Uncertain
Was the execution of the index test described in sufficient detail to permit replication of the test?	Yes, No, Uncertain
Was the execution of the reference standard described in sufficient detail to permit replication of the refererence standard?	Yes, No, Uncertain
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes, No, Uncertain
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes, No, Uncertain
Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	Yes, No, Uncertain
Were uninterpretable/intermediate tests results reported?	Yes, No, Uncertain
Were withdrawals from the study explained?	Yes, No, Uncertain
Overall Quality Assessment: (To be completed by Dr. Gutman or Dr. Bradley ONLY)	
ABSTRACTOR COMMENTS: If you would like to leave a comment pertaining to the information above indicate your name below:	

Appendix J. Detailed Description of PCA3 and Total PSA Interpretive Analysis

Introduction: Partial Verification Bias for Total PSA (tPSA)

Study design is an important inclusion criterion for this comparative review, because tPSA measurements, the main comparator to PCA3, are integral to men's decision-making regarding prostate biopsy. Men will be offered prostate biopsy based on the extent of tPSA elevations, suspicious findings on a digital rectal exam (DRE), a combination of the two or, less commonly, other risk factors such as family history or race. In order to obtain an unbiased estimate of diagnostic accuracy for tPSA at specific cut-offs, it is necessary that the identification of prostate cancer not be related to tPSA levels. This is a potential problem as most clinicians believe that higher tPSA levels are indicative of a higher likelihood for the presence of prostate cancer. Men are more likely to undergo prostate biopsy, if the tPSA is high (e.g., 10-20 ng/mL), rather than close to lower cut-offs used to define a positive tPSA screening test (e.g., 3-4 ng/mL). If a study reports results in which biopsy is tPSA-related, the sensitivity and specificity at select tPSA cut-offs will not be accurate. If those not accepting biopsy are considered missing, this is considered 'partial verification' bias. All studies included in the evidence review are opportunistic cohorts of men agreeing to biopsy and will likely be subject to this bias.

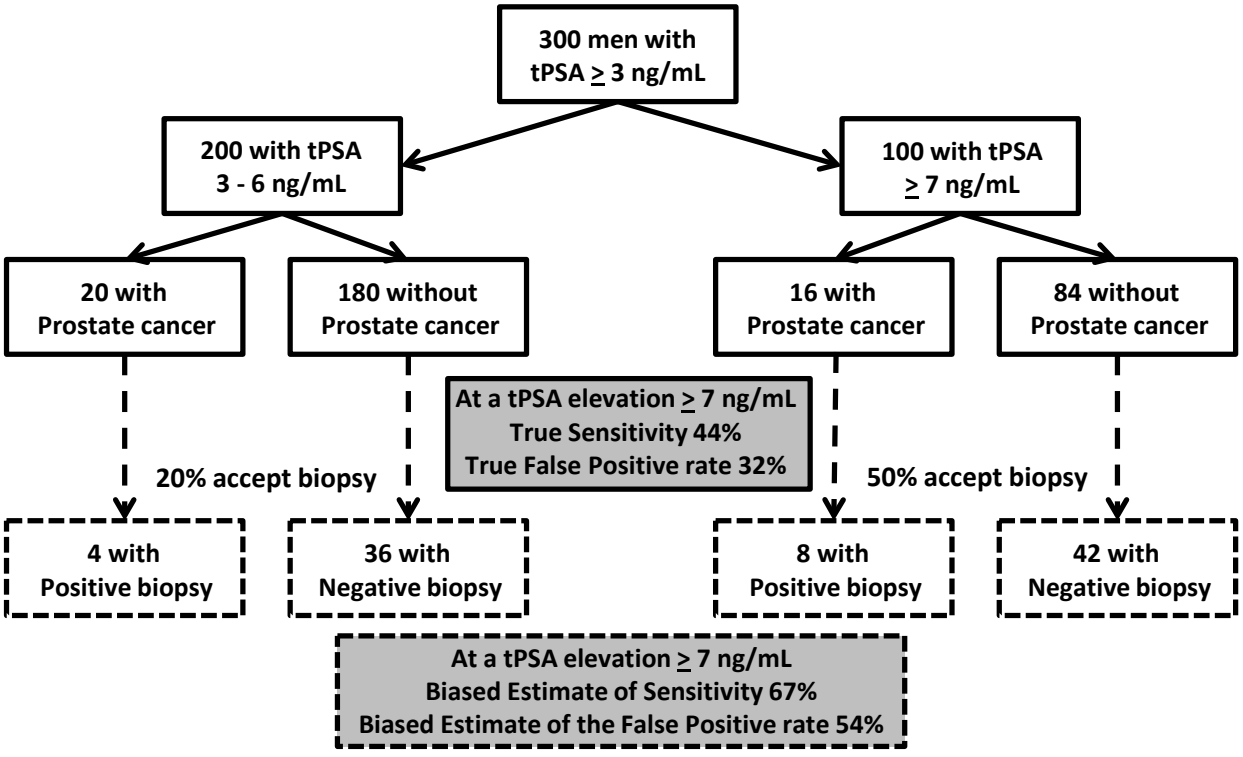


Figure J1. A schematic showing the impact of partial verification bias on the sensitivity and false positive rate for total PSA (tPSA) and prostate biopsy. A complete description is provided in the text.

An example of how partial verification bias occurs is demonstrated in Figure J1. These numbers are intended to be representative of clinical practice, but are mainly designed to

demonstrate partial verification bias in diagnostic studies of tPSA and prostate cancer. Among a cohort of as many as 3,000 men over 50 years of age, a subset of 300 men have been identified as having tPSA values of 3 ng/mL or higher. The remaining 2,700 men had low tPSA values and are not shown. For simplicity, consider all men with tPSA over three to be categorized into only two groups defined by the extent of tPSA elevation (3-6 ng/mL and ≥7 ng/mL). The prevalence of prostate cancer in the group with modestly elevated tPSA will be somewhat lower (10%) than the prevalence (16%) in the group with high tPSA elevations. At this point, the 'true' sensitivity and associated false positive rate (1-specificity) using an arbitrary tPSA cut-off of 7 ng/mL can be determined to be 44% (16/(16+20)) and 32% (84/(84+180)), respectively (shaded box with solid outline). The associated group likelihood ratio is 1.38 (44%/32%). These 'correct' results would have been obtained if all 300 men were to have had biopsies, assuming that biopsy is a perfect test for identify prostate cancer. A less biased result could also be obtained if, in addition to the subset having biopsies, all remaining men were followed longitudinally for several years to identify future cancer diagnoses.

In clinical practice, however, the uptake rate for prostate biopsy will depend on the group to which the men belong. In the group with lower tPSA levels, the uptake rate may be as low as 20%. In the group with higher tPSA levels, the uptake rate may be as high as 50%. These uptake rates, and results in men choosing biopsy are shown by the dashed arrows and boxes. Among the group of men choosing biopsy based loosely on the tPSA level, the sensitivity and false positive rate are now 67% (8/(8+4)) and 54% (42/(42+36)), respectively. The corresponding cumulative likelihood ratio is 1.24 (67/57).

Figure J1 shows that partial verification bias in this setting using a cut-off of 7 ng/mL tends to overestimate the sensitivity (67% versus 44%), as the men with cancer having with low tPSA are less likely to be identified (false negatives). The bias will also overestimate the false positive rate (54% versus 32%) as the men without cancer having low tPSA are also less likely to be identified (true negatives).

Literature Regarding Partial Verification Bias for tPSA elevations

An important paper on partial verification bias and diagnostic accuracy of tPSA addressed this population-based screening test.[1] That study focused on a data set of men having tPSA testing in which men with both normal and elevated tPSA levels were offered biopsies. This differs from the setting of the current review which focuses on using the extent of tPSA elevations among men already identified as being 'high risk' due to at least modest elevations of tPSA (e.g., >2.5 ng/mL, ≥3 ng/mL) and/or other factors already described. Punglia and colleagues reported that partial verification bias has two separate and important effects: 1) the performance of tPSA measurements is actually better than the reported performance subject to partial verification bias, and 2) the cut-off level at which select sensitivities and associated false positive rates occur are quite different in biased versus unbiased studies.[1] For example, they reported the area under the ROC curve for men under age 60 years to be 0.69 with partial verification bias, and 0.86 after accounting for the bias ($p<0.001$). In that same group, a tPSA cut-off of 4.1 ng/mL was associated with a sensitivity and false positive rate of about 43% and 22% with bias; 18% and 2% after adjustment. One limitation to that study was the use of a categorical uptake rate to model the presence of cancer in men without biopsies. This may have caused at least some of the difference in performance.

A later study[2] also adjusted for verification and found that the ROC curve was essentially unchanged (AUC 0.682 among verified, and 0.678 after accounting for partial verification).

However, a relatively high proportion (64%) of the population was biopsied. We verified by in-house modeling that in the setting of the current review, it is likely that the ROC will not be substantially impacted (modeled uptake proportional to tPSA, ranging from 5% to 50%, with 30% overall verification). Several additional studies have assumed that a single cut-off will be used, and focus only on correcting the estimates of sensitivity and specificity at that cut-off, and do not consider the issue of whether the biased estimates of sensitivity and specificity falls on the true ROC curve or not.[3-7]

There are two important factors that appear to influence the extent of partial verification bias:

1. the differential rates of cancer (positive prostate biopsy) in the group below the tPSA cut-off compared to the group above the cut-off. If the rates of cancer are the same in the two groups, no partial verification bias will occur. The greater the relative difference (rate ratio) between groups, the larger effect the partial verification bias will have.
2. the differential biopsy uptake rates in the two groups. If the biopsy uptake rates are the same, there will be no partial verification bias, even if the rates of cancer are different. Analogously, if there is no difference in the cancer rates, there will be no partial verification bias even if the uptake rates differ.

The setting of the current evidence review differs from that explored by Puglia and colleagues[1] in two important ways. First, the differential rate of cancer in studies relevant to this review (all having at least modest elevations) is likely to be considerably less. This suggests that the partial verification bias might have less impact in the current setting. Second, the biopsy uptake across the more limited range of tPSA in this review/setting is likely to be more similar. Lastly, the overall uptake rate will also be higher, as all of the men in our setting would all have an elevated tPSA (and/or positive DRE or other factor). All of these differences would tend to reduce the effect of partial verification bias. Of most importance is our finding that the ROC curves (and, by extension, AUC and sensitivity/specificity estimates) are likely to not be impacted by partial verification bias for tPSA measurements, confirming a previous finding[2].

Review of Methodology Specific to the PCA3/tPSA Comparison

The aims of the following interpretative analyses are to estimate the unbiased performance of tPSA and PCA3 to identify ‘at risk’ men who will have a negative, or positive, biopsy. The analysis will be anchored by two important findings. First, the ROC curves for tPSA (and for PCA3) are not influenced by the partial verification bias. Thus, any set of parameters (distribution descriptors such as means and standard deviations for PCA3 and comparators in both biopsy negative and positive men) we might generate, would have to fit that data. Second, the literature contains sufficient information to estimate the needed parameters, but those estimates will be subject to the partial verification bias for tPSA and related markers, but not for PCA3. If the reported tPSA parameters can be ‘unbiased’ and fitted to the relevant ROC curve, then a direct comparison between tPSA and PCA3 at selected cut-off levels could be made. Based on published data, the distributions of these markers are likely to be Gaussian, after a logarithmic transformation. To simplify the analysis, we have also chosen to force the logarithmic standard deviations within a study to be the same for each marker. This was accomplished by an un-weighted pooling of the estimated variances. According to the literature, this seems to be a reasonable assumption. The secondary aims of these analyses are to provide improved templates for more reliably exploring the comparison of prostate cancer markers and assist in providing methods to more fully inform decision-making by men and their health care providers.

Interpretative Analysis for PCA3

Figure J2 shows how overlapping distributions of PCA3 can be used to generate sensitivity and false positive rates (1-specificity). The baseline data used to estimate the median and logarithmic SDs are provided in the body of the evidence review (Table 11). Eight studies[8-15] reported sufficient data for this analysis, after excluding the two studies[14, 15] that focused on the grey zone of tPSA. Although the reported median values are somewhat variable, the PCA3 scores are about 2.2 (median ratio) times as high in those with a positive versus a negative biopsy. These results are not subject to partial verification bias because tPSA and PCA3 levels are essentially independent markers (Evidence Review, Table 15). There is more consensus on the pooled logarithmic SDs being at about 0.420. The median PCA3 score among those with a negative biopsy was 21. Using the 2.2 multiplier, a reasonable expected median PCA3 score among those with a positive biopsy would be around 46 (2.2 * 21). These population parameters fit the consensus ROC data well, as shown in the next paragraph and Figure J3.

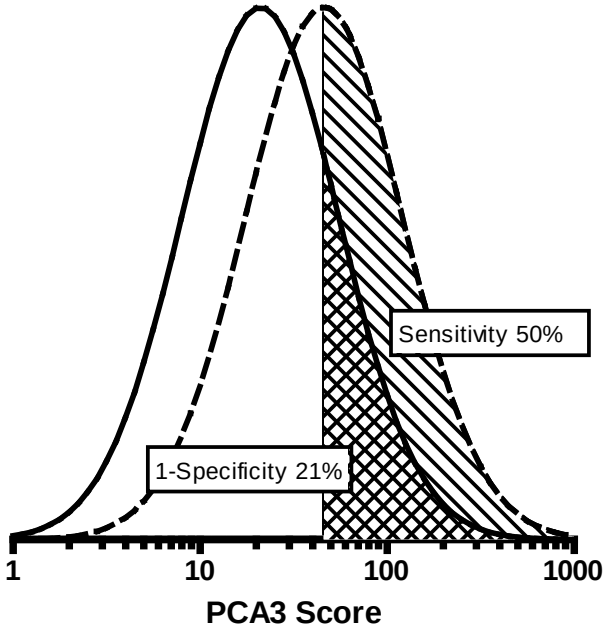


Figure J2. Overlapping distributions of PCA3 scores in men with negative or positive biopsies.
The overlapping distributions are described by log Gaussian parameters (1.322, 0.420: solid line, negative and 1.663, 0.420; dashed line, positive). A vertical line has been drawn at 46 to demonstrate how these data can be used to generate sensitivity (50%, hatched area) and the false positive rate (21%, cross-hatched area). Moving the line left and right will generate all points on the corresponding ROC curve for PCA3 (Figure J3).

Figure J3 shows seven points on the ROC curve for PCA3 derived from the literature review (Evidence review, Table 13). These points correspond to false positive rates of 20% through 80%, in 10% intervals. The dashed line indicates a 'useless' test as a reference (sensitivity equal to the false positive rate). The solid line has been derived from the overlapping distributions shown in Figure J2, and fit well. It is not possible to formally test the fit, as most of the original data were obtained from the published figures, rather than from original data. This model allows the derivation of expected sensitivity and false positive rates across the range of specific cut-off values, and provides an estimate of what PCA3 score is associated with that test performance. According to the relevant studies, PCA3 scores in men with negative (and positive) biopsies vary between reports, so the modeled PCA3 cut-offs may not be appropriate in all settings. For this reason, the results of this analysis should be used only as a guide. This gap in knowledge could be addressed in future research by examining normalizing functions such as multiple of the median PCA3 score, where the median level is determined in a healthy male population without prostate cancer.

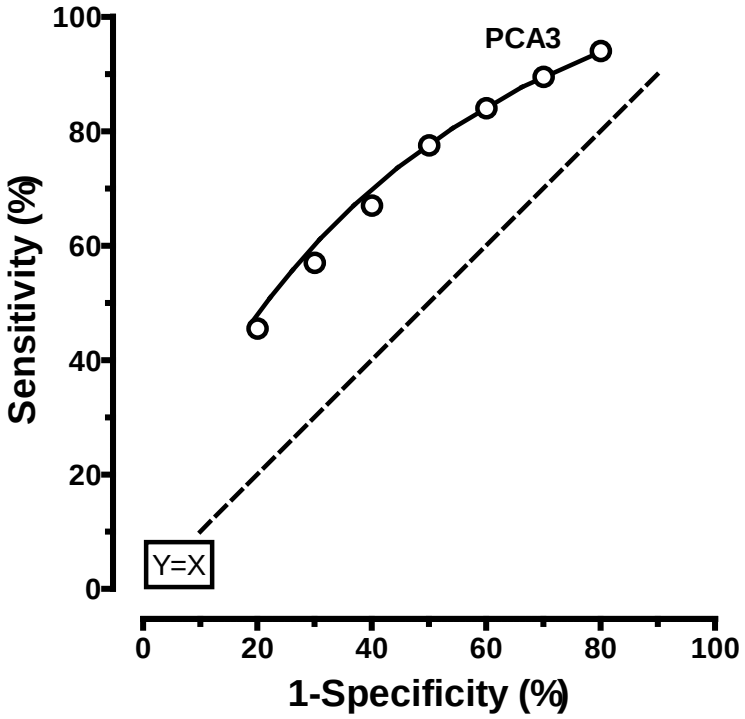


Figure J3. A fitted ROC curve for PCA3 scores.
The figure shows that the overlapping distributions provided in Figure J2 Adequately fit the observed ROC data from the literature (open circles). Simple Gaussian modeling can now be used to generate sensitivity at false positive rates between 20% and 80%.

Interpretative Analysis for tPSA

Similar data were available to create a tPSA model, but with the caveat that partial verification bias must be accounted for in determining the parameters. The tPSA literature was more consistent than PCA3 in the reported median and logarithmic SDs for tPSA (Evidence Review, Table 9). For the same six studies[8-13], the median tPSA in those with a negative biopsy was 6.3 ng/mL. The median ratio between the tPSA in positive versus negative biopsies was 1.30. Thus, the expectation is that the tPSA level would be about 8.2 ng/mL (6.3 x 1.3) in men with positive biopsies. The consensus logarithmic SD was about 0.31. Unfortunately, partial verification bias will inflate the median value, as lower tPSA measurements in men both with and without cancer will be under-represented. The bias will also tend to shrink the logarithmic standard deviation. In order to estimate how much lower and broader to make the distributions, we generated log Gaussian distributions and subjected them to partial verification bias to study the effect. In one of these models, the uptake rate was estimated to be the tPSA value divided by 20, with an upper limit of 0.5 (50% uptake). For example, men with a tPSA of 8 ng/mL would have an uptake rate of 40% (8/20) while those with a tPSA of 4 ng/mL would have a 20% uptake rate. We found the variance increased by about 40%, while the median value was reduced by about 35%. When we applied these correction factors to the consensus reported medians and logarithmic SD described above, slight adjustments in the reduction in median values were still needed to more closely match the reported tPSA ROC curve. The final parameters are shown in Figure J5. These data must be viewed as approximate. However, knowing that the necessary data to produce an observed and unbiased estimate is likely to be unattainable, it may be preferable to reporting that the analysis cannot proceed due to the existence of partial verification bias.

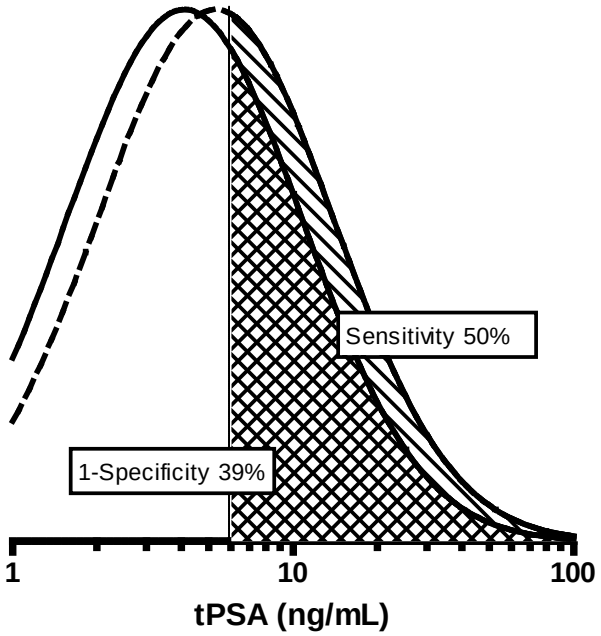


Figure J4. Overlapping distributions of tPSA measurements in men with negative or positive biopsies, after accounting for partial verification bias. The overlapping distributions are described by log Gaussian distributions (0.613, 0.420: solid line, negative biopsy and 0.724, 0.420; dashed line, positive biopsy). A cut-off (vertical line) has been drawn at 5.3 ng/mL to demonstrate how these data can be used to generate sensitivity (50%, hatched area) and the false positive rate (39%, cross-hatched area). Moving the line left and right will generate all points on the corresponding ROC curve for tPSA.

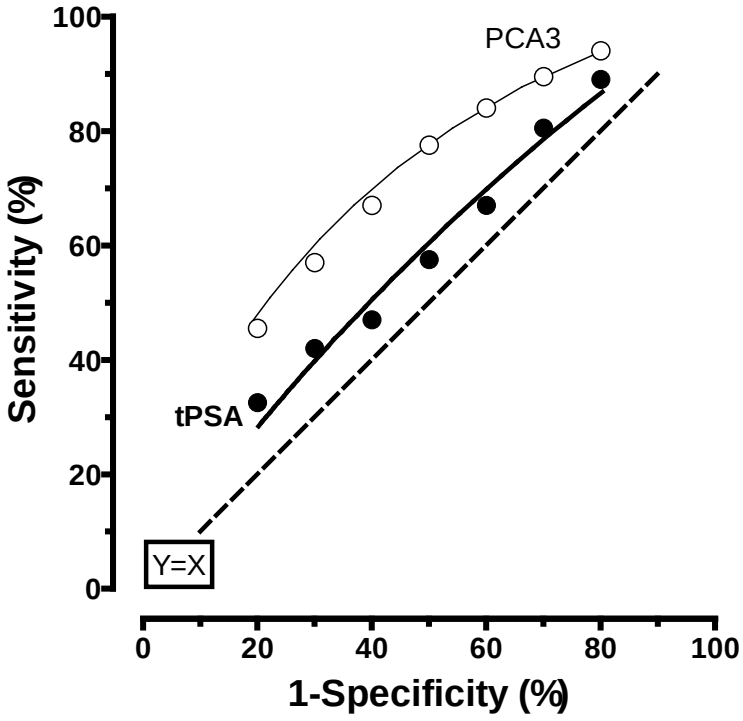


Figure J5. A fitted ROC curve for tPSA measurements, after accounting for partial verification bias. The figure shows that the overlapping distributions provided in Figure J4 Adequately fit the observed ROC data from the literature (open circles). Simple Gaussian modeling can now be used to generate sensitivity at false positive rates between 20% and 80%. The PCA3 curve from Figure J3 has been included for comparison (thin line, open circles).

Figure J5 shows seven points on the ROC curve for tPSA derived from the literature review. Only data from the six studies[8-13] that did not restrict tPSA measurements to the 'grey zone' were included (Evidence Review, Table 11). These points will be essentially unchanged whether or not partial verification bias is accounted for, and correspond to false

positive rates of 20% through 80%. The dashed line indicates a 'useless' test as a reference (sensitivity equal to the false positive rate). The solid line has been derived from the overlapping distributions show in Figure J4, and fits well.

Interpretative Analysis Comparing PCA3 and tPSA Performance

These parameter sets for PCA3 scores and tPSA elevations can now be used to compare their performance in identifying men who will have a positive versus a negative biopsy. The following analyses rely only on the parameters presented in the legends to Figures J3 and J5 along with three selected rates of cancer in the population (5%, 10% and 15%). The reported rates from the reviewed literature ranged between 10% and 30%. However, due to partial verification bias, the actual prostate cancer rates are likely to be somewhat lower.

Table J1 presents a series of PCA3 score cut-off levels with accompanying false positive rates (1-specificity) and sensitivity (detection) rates. For example, using the common PCA3 cut-off score of 35 and higher, 29.9% of men without cancer would still undergo biopsy. Alternatively, this can be looked at as avoiding an unneeded biopsy in70.1% of men without cancer. The trade-off to this reduction in biopsy is that the detection rate is only 61.1%, indicating that this would result in 38.9% of prostate cancers in the population being missed. The last two columns include the cumulative likelihood ratio (sensitivity / false positive rate) and the individual likelihood ratio (ratio of the heights of the two curves at the given value). For example, using the same score of 35 as the cut-off, the screen positive men will have 2.0 times the risk of prostate cancer compared to the entire cohort of men identified with elevated tPSA levels and/or positive DRE or other findings. For men with a score of exactly 35, their risk is increased by only 16% (LR=1.16) over the cohort as a whole.

The bottom half of Table J1 presents the data in the same way, but for tPSA measurements. For example, 34.7% of the men who would have a negative biopsy would still have a tPSA of six ng/mL or higher (false positive rate). Correspondingly, about 44.9% of the men with cancer would have levels at or above six ng/mL. (It is important to remember, this 44.9% is a conditional proportion of men with cancer among the cohort of men over age 50 identified with an initially elevated tPSA and/or with other markers. It is not the proportion of all men over age 50 with prostate cancer.) The cumulative likelihood indicates an increase in risk of 30% (LR=1.3), while an individual man with a tPSA of 6 ng/mL would have little change in risk (LR=1.07). To allow for direct comparisons, the range of cut-offs shown for the two markers cover approximately the same proportion of the overlapping distributions. The format of Table J1 makes comparisons between the two markers difficult, as neither the sensitivity nor the false positive rate is being held constant.

Table J2 compares the performance of PCA3 scores and tPSA measurement to identify men who would (or would not) have a positive biopsy. The top half of the Table J2 holds the false positive rate constant, while the bottom half holds the sensitivity constant. In the last column is the difference between the two estimates (PCA3 – tPSA). When comparing the sensitivities (top half) this column contains the improvement in prostate cancer detection at a fixed false positive rate. When comparing the false positive rates, it contains the reduction in unnecessary biopsies.

Table J1. Sensitivity, false positive rates and likelihood ratios for PCA3 and tPSA elevations.

Table J1 A. PCA3

PCA3 Score	1-Specificity (FPR)	Saved Biopsy	Sensitivity (DR)	PCa Missed	Population LR	Individual LR
≥10	77.9%	22%	94.3%	6%	1.2	0.39
≥15	63.6%	36%	87.7%	12%	1.4	0.54
≥20	52.0%	48%	80.5%	19%	1.5	0.69
≥25	42.8%	57%	73.6%	26%	1.7	0.83
≥30	35.6%	64%	67.1%	33%	1.9	0.97
≥35	29.9%	70%	61.1%	39%	2.0	1.10
≥40	25.3%	75%	55.7%	44%	2.2	1.24
≥45	21.5%	78%	50.9%	49%	2.4	1.36

Table J1 B. tPSA

tPSA (ng/mL)	1-Specificity (FPR)	Saved Biopsy	Sensitivity (DR)	PCa Missed	Population LR	Individual LR
≥2	77.1%	23%	84.3%	16%	1.1	0.79
≥3	62.7%	37%	72.2%	28%	1.2	0.89
≥4	51.0%	49%	61.4%	39%	1.2	0.96
≥5	41.9%	58%	52.4%	48%	1.3	1.02
≥6	34.7%	65%	44.9%	55%	1.3	1.07
≥7	29.0%	71%	38.7%	61%	1.3	1.12
≥8	24.5%	76%	33.5%	66%	1.4	1.16
≥9	20.8%	79%	29.2%	71%	1.4	1.20

LR=Likelihood ratio; PCa=prostate cancer; DR=detection rate; FPR=false positive rate (1-specificity)

Table J2. A comparison of PCA3 scores and tPSA elevations to identify men with prostate cancer with either the false positive rate, or the sensitivity held constant.

Table J2 A. PCA3 Measurements

1-Specificity (FPR) Held constant	PCA3 Cut-off (Score)	PCA3 Sensitivity (DR)	tPSA Cut-off (ng/mL)	tPSA Sensitivity (DR)	Result: Improvement in PCa Detection
80%	≥9.3	95.1%	≥1.8	86.8%	8.3%
70%	≥12.6	91.0%	≥2.5	78.1%	12.9%
60%	≥16.4	88.7%	≥3.2	72.2%	16.5%
50%	≥21.0	79.1%	≥4.1	60.5%	18.6%
40%	≥26.8	71.2%	≥5.2	50.8%	20.4%
30%	≥34.9	61.2%	≥6.8	39.8%	21.4%
20%	≥47.4	48.8%	≥9.3	28.0%	20.8%

Table J2 B. tPSA Measurements

Sensitivity (DR) Held constant	PCA3 Cut-off (Score)	PCA3 1-Specificity (FPR)	tPSA Cut-off (ng/mL)	tPSA 1-Specificity (FPR)	Result: Reduction in Biopsies
95%	≥9.3	80.0%	≥1.1	89.8%	9.8%
90%	≥13.3	68.2%	≥1.5	85.1%	16.9%
85%	≥16.9	58.9%	≥1.9	78.7%	19.8%
80%	≥20.4	50.2%	≥2.3	72.5%	22.3%
70%	≥27.7	38.7%	≥3.2	60.1%	21.4%
60%	≥36.0	28.9%	≥4.1	50.0%	21.1%
50%	≥46.0	20.9%	≥5.3	39.5%	18.6%

PCa=prostate cancer; DR=detection rate, FPR=false positive rate (1-specificity)

As an example, assume you would like to choose a cut-off so that no more than 10% of the existing cancers would be missed (*i.e.,* false negative on the PCA3 or tPSA tests). Choose the highlighted row with 90% sensitivity (Table J2). It shows the false positive rate would be 68% using PCA3 scores, but 85% using tPSA measurements. This means that the same proportion (90%) of cases might be detectable while performing 17% fewer biopsies were PCA3 to be used instead of tPSA elevations.

These analyses do not account for the prevalence of cancer in the cohort, so cannot determine the positive or negative predictive values. Figure J6 shows how these data can be used with cancer prevalence to determine predictive values. It also shows how the cohort of interest is limited to the population of men who have elevated tPSA/positive DRE. Assume the prevalence of cancer is 5%, and the PCA3 related sensitivity is chosen to be 90%. Based on Table J2, this implies a PCA3 cut-off score of 13.3 and will have a false positive rate of 68.2%. (As an aside, the analysis is more confident in the performance estimates of 90% sensitivity and corresponding false positive rate than that those rates occur at a score of 13.8.) Depending on the situation, the relevant PCA3 score may be higher or lower, depending on factors that have not yet been clarified.)

The flowchart (Figure J6) first assumes that a cohort of 3,000 men is subject to prostate cancer screening and 10% are found to be 'at risk' due to elevated tPSA, positive DRE and/or other factors. Only these 300 are relevant to the current evidence review. The PCA3 score is then applied to this subset of 300 men, 30 of whom have prostate cancer (10%). Using the example from the previous paragraph, assume a 90% sensitivity for PCA3 testing (cut-off score of 13.3). Among the 30 men with cancer, the PCA3 score will be 13.3 or higher in 27 (90%). Among the 270 men without cancer, PCA3 will be similarly elevated in 184 (68.2%). Among the 211 with a positive PCA3 (27 TP + 184 FP), the positive predictive value (PPV) is 13% (27/211). Another way to express the PPV is via odds. The PPV of 13% is equivalent to odds of 1:6.8 (this could be read as odds of 1 to 6.8 or as a proportion of 1 in 7.8).

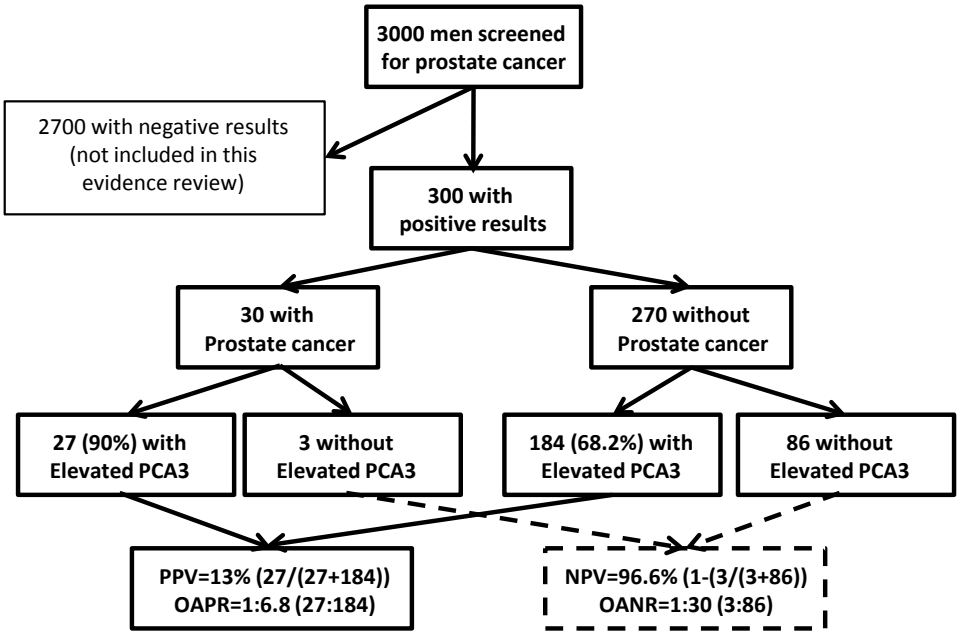


Figure J6. Flowchart showing the cohort of interest and how the positive and negative predictive values are computed. The flowchart shows the performance of PCA3 testing among 300 men with positive prostate screening test results. More information is provided in the text.

This is also called the odds of being affected given a positive result (OAPR). Among the 89 men with a PCA3 score under 13.3, 86 men will not have cancer, resulting in a negative predictive value (NPV) of 96.6%. Another way to understand the NPV is to create odds of having cancer among those with a negative test result. This is called the odds of being affected given a negative result (OANR); in this example, 1:30.

Table J3 provides a summary of positive and negative predictive values for both PCA3 scores and tPSA at seven different sensitivities ranging from 50% to 95%, each with three different cancer rates (5%, 10% and 15%). Within each of the groups with the same sensitivity, the PPV increase while the NPV decreases. Between groups, increasing sensitivity is associated with higher PPV and lower NPV rates, at the same cancer rate. When comparing across the table (PCA3 versus tPSA), both the PPV and NPV are slightly lower.

Up to this point, all tables and figures were designed to assist health care providers understand each test's performance, evaluate the trade-offs at selected cut-off levels, and compare performance between tests. This is of less interest to an individual male who does not have a PCA3 (or tPSA) level at or above a cut-off level, but instead has a 'patient –specific' measurement that could be interpreted for that individual. Table J4 provides this information at select PCA3 and tPSA levels. There are, however, too many possibilities to provide a complete listing, but this type of information could easily be part of a computerized report. Such information could also be tailored to include other relevant risk factors such as family history of prostate cancer or race. The table allows for prior risks of between 5% and 30%. As an example, consider a male with a relatively low PCA3 score of 10. Where his prior risk to be low at 5% (*e.g.*, 50 y.o. white male, no family history and his initial screening test results to be only slightly elevated with a negative DRE), then the PCA 3 score would reduce his risk to 1:106. However, if his PCA3 score more elevated, at 45, his risk would be slightly increased to 1:14. In general, a stronger marker is more able to modify the initial risks. This can be seen by scanning the first column of risks. PCA3 risks vary more than do the corresponding tPSA-related risks, even though the ranges of values are similar (footnote, Table J4).

Table J3. Positive and negative predictive values for PCA3 and tPSA testing at seven selected test sensitivities, each with three different rates of prostate cancer

Sensitivity (DR)	PCa Rate (%)	PCA3 (Score) FPR	PCA3 (Score) PPV (OAPR)	PCA3 (Score) NPV (OANR)	tPSA (ng/mL) FPR	tPSA (ng/mL) PPV (OAPR)	tPSA (ng/mL) NPV (OANR)
95	5	80.0	5.9% (1:15.8)	98.7% (1:76.0)	89.8	5.3% (1:17.9)	97.5% (1:38.8)
	10	80.0	11.9% (1: 7.4)	97.3% (1:36.0)	89.8	10.6% (1: 8.5)	94.8% (1:18.4)
	15	80.0	17.8% (1: 4.6)	95.8% (1:22.7)	89.8	15.9% (1: 5.3)	92.0% (1:11.6)
90	5	68.2	6.6% (1:14.2)	98.4% (1:60.4)	85.1	5.3% (1:17.9)	96.6% (1:28.3)
	10	68.2	13.2% (1: 6.6)	96.6% (1:28.6)	85.1	10.6% (1: 8.5)	93.1% (1:13.4)
	15	68.2	19.8% (1: 4.1)	94.7% (1:18.0)	85.1	15.9% (1: 5.3)	89.4% (1: 8.4)
85	5	58.9	7.2% (1:12.9)	98.1% (1:52.1)	78.7	5.4% (1:17.5)	96.4% (1:27.0)
	10	58.9	14.4% (1: 5.9)	96.1% (1:24.7)	78.7	10.8% (1: 8.3)	92.7% (1:12.8)
	15	58.9	21.6% (1: 3.6)	93.9% (1:15.5)	78.7	16.2% (1: 5.2)	88.9% (1: 8.0)
80	5	50.2	8.0% (1:11.6)	97.9% (1:47.3)	72.5	5.5% (1:17.1)	96.3% (1:26.1)
	10	50.2	15.9% (1: 5.3)	95.7% (1:22.4)	72.5	11.0% (1: 8.1)	92.5% (1:12.4)
	15	50.2	23.9% (1: 3.2)	93.4% (1:14.1)	72.5	16.6% (1: 5.0)	88.6% (1: 7.8)
70	5	38.7	9.0% (1:10.1)	97.5% (1:38.8)	60.1	5.8% (1:16.2)	96.2% (1:25.3)
	10	38.7	18.1% (1: 4.5)	94.8% (1:18.4)	60.1	11.6% (1: 7.6)	92.3% (1:12.0)
	15	38.7	27.1% (1: 2.7)	92.1% (1:11.6)	60.1	17.5% (1: 4.7)	88.3% (1: 7.5)
60	5	28.9	10.4% (1: 8.6)	97.1% (1:33.8)	50.0	6.0% (1:15.7)	96.0% (1:23.8)
	10	28.9	20.8% (1: 3.8)	94.1% (1:16.0)	50.0	12.% (1: 7.3)	91.8% (1:11.3)
	15	28.9	31.1% (1: 2.2)	91.0% (1:10.1)	50.0	18.% (1: 4.6)	87.6% (1: 7.1)
50	5	20.9	12.0% (1: 7.4)	96.8% (1:30.1)	39.5	6.3% (1:14.8)	95.8% (1:23.0)
	10	20.9	23.9% (1: 3.2)	93.4% (1:14.2)	39.5	12.7% (1: 6.9)	91.6% (1:10.9)
	15	20.9	35.9% (1: 1.8)	90.0% (1: 9.0)	39.5	19.0% (1: 4.3)	87.3% (1: 6.9)

DR = detection rate, PCa = prostate cancer, FPR = false positive rate (1-specificity), PPV = positive predictive value, OAPR = odds of being affected given a positive results, NPV = negative predictive value, OANR = odds of being affected given a negative result

Table J4. Patient-specific prostate cancer risks by an individual's prior risk of prostate cancer (1:N) and by PCA3 score or tPSA elevation.

Table J4 A. PCA3 Score

PCA3 Score[a]	1:19 (5%)	1:9 (10%)	1:5.7 (15%)	1:4 (20%)	1:3 (25%)	1:2.3 (30%)
5	106	50	32	22	17	13
10	88	42	26	19	14	11
15	49	23	15	10	7.8	6.0
20	35	17	10	7.4	5.5	4.3
25	27	13	8.2	5.8	4.3	3.4
30	23	11	6.8	4.8	3.6	2.8
35	20	9.3	5.8	4.1	3.1	2.4
40	17	8.1	5.1	3.6	2.7	2.1
45	15	7.3	4.6	3.2	2.4	1.9
50	14	6.6	4.2	2.9	2.2	1.7

Table J4 B. tPSA Elevations

tPSA (ng/mL)[1]	1:19 (5%)	1:9 (10%)	1:5.7 (15%)	1:4 (20%)	1:3 (25%)	1:2.3 (30%)
3	21	10	6	4.5	3.4	2.6
4	20	9	5.9	4.2	3.1	2.4
5	19	9	5.6	3.9	2.9	2.3
6	18	8	5.3	3.7	2.8	2.2
7	17	8.0	5.1	3.6	2.7	2.1
8	16	7.8	4.9	3.4	2.6	2.0
9	16	7.5	4.7	3.3	2.5	1.9
10	15	7.3	4.6	3.2	2.4	1.9
11	15	7.1	4.5	3.2	2.4	1.8
12	15	6.9	4.4	3.1	2.3	1.8

[1] Approximate range: mean of biopsy negative minus 1 SD to mean of biopsy positive plus 0.5 SD.

References

1. Punglia RS, D'Amico AV, Catalona WJ, et al. Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. N Engl J Med. 2003 Jul 24;349(4):335-42. PMID: 12878740.
2. Thompson IM, Ankerst DP, Chi C, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. JAMA. 2005 Jul 6;294(1):66-70. PMID: 15998892.
3. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983 Mar;39(1):207-15. PMID: 6871349.
4. de Groot JA, Janssen KJ, Zwinderman AH, et al. Correcting for partial verification bias: a comparison of methods. Annals of epidemiology. 2011 Feb;21(2):139-48. PMID: 21109454.
5. de Groot JA, Janssen KJ, Zwinderman AH, et al. Multiple imputation to correct for partial verification bias revisited. Stat Med. 2008 Dec 10;27(28):5880-9. PMID: 18752256.
6. Gupta ARoehrborn CG. Verification and incorporation biases in studies assessing screening tests: prostate-specific antigen as an example. Urology. 2004 Jul;64(1):106-11. PMID: 15245945.
7. Rosman ASKorsten MA. Effect of verification bias on the sensitivity of fecal occult blood testing: a meta-analysis. Journal of general internal medicine. 2010 Nov;25(11):1211-21. PMID: 20499198.
8. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008 Oct;180(4):1303-8; discussion 08. PMID: 18707724.
9. Auprich M, Augustin H, Budaus L, et al. A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. BJU Int. 2012 Jun;109(11):1627-35. PMID: 21939492.
10. Bollito E, De Luca S, Cicilano M, et al. Prostate cancer gene 3 urine assay cutoff in diagnosis of prostate cancer: a validation study on an Italian patient population undergoing first and repeat biopsy. Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology. 2012 Apr;34(2):96-104. PMID: 22611765.
11. Hessels D, van Gils MP, van Hooij O, et al. Predictive value of PCA3 in urinary sediments in determining clinico-pathological characteristics of prostate cancer. Prostate. 2010 Jan 1;70(1):10-6. PMID: 19708043.
12. Nyberg M, Ulmert D, Lindgren A, et al. PCA3 as a diagnostic marker for prostate cancer: a validation study on a Swedish patient population. Scand J Urol Nephrol. 2010 Dec;44(6):378-83. PMID: 20961267.
13. Ochiai A, Okihara K, Kamoi K, et al. Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. Int J Urol. 2011 Mar;18(3):200-5. PMID: 21332814.
14. Ferro M, Bruzzese D, Perdona S, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta. 2012 Aug 16;413(15-16):1274-8. PMID: 22542564.
15. Perdona S, Cavadas V, Di Lorenzo G, et al. Prostate cancer detection in the "grey area" of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. European urology. 2011 Jan;59(1):81-7. PMID: 20947244.
16. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol. 2008 Apr;179(4):1587-92. PMID: 18295257.
17. Goode RR, Marshall SJ, Duff M, et al. Use of PCA3 in detecting prostate cancer in initial and repeat prostate biopsy patients. Prostate. 2012 May 14; PMID: 22585386.
18. Rigau M, Morote J, Mir MC, et al. PSGR and PCA3 as biomarkers for the detection of prostate cancer in urine. Prostate. 2010 Dec 1;70(16):1760-7. PMID: 20672322.
19. Schilling D, Hennenlotter J, Munz M, et al. Interpretation of the prostate cancer gene 3 in reference to the individual clinical background: implications for daily practice. Urol Int. 2010 85(2):159-65. PMID: 20424427.

20. Wu AK, Reese AC, Cooperberg MR, et al. Utility of PCA3 in patients undergoing repeat biopsy for prostate cancer. Prostate Cancer Prostatic Dis. 2012 Mar;15(1):100-5. PMID: 22042252.

Appendix K. Summary of the Remaining Combined Analyses for KQ 1 and KQ 2

Comparator: Percent Free serum PSA (%fPSA)

Intermediate Outcome: Diagnostic Accuracy

Key Points -The extent of %fPSA elevations were compared with PCA3 scores to determine their diagnostic accuracy to predict prostate biopsy results (cancer / no cancer). Six of eight included studies (Table K1) enrolled only men with tPSA elevations in the grey zone (usually 2 to 10 ng/mL).

- Area under the curve (AUC). Five studies[1-5] reported AUC results for PCA3 and %fPSA (Table K2). Three studies[1, 4, 5]focused on the 'grey zone' of tPSA values. Two 'grey zone' studies[1, 5]found PCA3 to be better by 12 percent to 17 percent, while two studies that enrolled men with all levels of tPSA[2, 3] found PCA3 and %fPSA to be similar (-3% and +6%). The third 'grey zone' study[4] did not report the actual AUC for %fPSA, but did report the PCA3 AUC to be significantly higher.
- Reported medians and standard deviations (SD). Three studies[2, 3, 6] provided some data for analysis (Table K3). One study in the 'grey zone'[6] found PCA3 to be better (+0.86 z). Another[2] found %fPSA to be better (-0.91 z) and the third[3] found them to be nearly equivalent (+0.15 z). Both of these latter studies finding smaller effects were not restricted to the 'grey zone'.
- Performance at a PCA3 cut-off score of 35. Only one 'grey zone' study[7] (Table K4) reported the sensitivity and specificity of PCA3 at this cut-off that could be compared with the %fPSA sensitivity. PCA3 sensitivity was higher by 17 percent.
- ROC Curves - Sensitivity/Specificity. Five studies[1, 3-5, 7] (Table K5) provided a ROC curve, or data representing an ROC curve, for both markers. Two datasets were available from one study[3], providing results for both initial and repeat biopsy. At all specificities noted, the median (PCA3 - %fPSA sensitivities) showed PCA3 to be slightly better. However, four of the six datasets were from studies focusing on the 'grey zone'.[1, 4, 5, 7]
- Regression analysis. Two studies[6, 7] (Table K6) provided sufficient data to apply the respective regression coefficients to create a relative odds ratio (OR) between the 25th and 75th centiles of the two distributions. Both studies had ORs showing PCA3 to perform better (2.88, 1.70), and both focused on the 'grey zone' of tPSA.

Study Characteristics

A total of seven studies[1-7] reported PCA3 and %fPSA comparisons that could be used in one or more of the matched analyses (Table K1). The table is sorted by the number of enrolled men. Five of the seven studies[1, 4-7] focused on the 'grey zone' of tPSA elevations. Given that there are likely to be correlations between tPSA and %fPSA, these studies might be expected to show a greater difference between PCA3 and %fPSA performance than studies in a more general population of men with tPSA elevations. Six[1, 3-7] of the seven studies were consistent in finding the matched PCA3 estimate better at identifying positive prostate biopsies than the corresponding %fPSA estimate. The exception was a small study[2] that enrolled men without regard to their tPSA measurements (*i.e.*, not 'grey zone'). Overall, the PCA3 and %fPSA

comparative findings were likely generalizable to only men in the 'grey zone' of tPSA levels at about 2 to 10 ng/mL (Ferro[4] 2 to 20 ng/mL).

Table K1. Summary results for the eight available analytic comparisons[a] of PCA3 versus %fPSA in matched populations of men having prostate biopsies

Study/Author[b]	Year	Number	AUC	Mean/SD	PCA3>35	Sens/Spec	Reg
Auprich[2]	2011	127	0.0270	-0.91	-	-	-
Ferro[4]	2012	151	NR[c]	-	-	19%	-
Perdona[6]	2011	218	-	0.86	-	-	2.88
Ploussard[5]	2010	301	0.1170	-	-	18%	-
de la Taille[7]	2011	516	-	-	17%	23%	1.70
Aubin[1]	2010	1072	0.1740	-	-	3%	-
Bollito[3]	2012	1246	0.0580	0.15	-	1% (initial) -9% (repeat)	-
All		**3705**					

[a] AUC = area under the curve for PCA3 minus the AUC of tPSA; Mean/DS = difference in separation between PCA3 scores and tPSA results, when expressed as z-scores; PCA3>35 = difference of the PCA3 minus the tPSA sensitivities at the specificity found for a PCA3 cut-off of 35; Sens/Spec = difference between PCA3 and tPSA sensitivity at a specificity of 50%; Reg = relative change in the PCA3 ORs (between the 25th and 75th centiles) and the corresponding tPSA ORs. The corresponding full analyses resulting in these summaries can be found on the following pages.

[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

[c] The AUC for %fPSA was not reported, but the PCA3 AUC was significantly higher (p=-0.036) than the unreported %fPSA AUC.

PCA3 and %fPSA: Area Under the Curve

Five studies[1-5] reported data on AUCs, and are presented in Table K2 sorted by study size. Three[1, 4, 5] of the five studies recruited only men in the 'grey zone' of tPSA (two of these found the AUC of PCA3 to be more than 10% higher, and the other did not report values, but did report that the AUC of PCA3 was significantly higher). In the two studies in the general population of PSA positive men[2, 3], one found %fPSA to be slightly better (-0.027), and the other found PCA3 to be slightly better (0.058). Due to the variability in the study populations and the limited number of studies, no further analyses are presented.

Table K2. Comparing performance of PCA3 levels and percent free PSA (%fPSA) measurements in matched studies via AUC analysis to correctly diagnose prostate cancer, as defined by a positive biopsy

Author[a]	Year	Number	Initial Bx	PCA3 AUC	%fPSA AUC	Difference[b]	P-value[c]
Auprich[2]	2011	127	0%	0.7030	0.7300	-0.0270	
Ferro[4]	2012	151	100%	0.7100	NR	-	0.036
Ploussard[5]	2010	301	0%	0.6880	0.5710	0.1170	-
Aubin[1]	2010	1072	0%	0.6930	0.5190	0.1740	0.04
Bollito[2]	2012	1246	59%	0.6780	0.6200	0.0580	
All		**2897**					

Bx = biopsy, AUC = area under the curve, '-' = no value reported, NR = not reported

[a] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

[b] AUC (PCA3) – AUC (tPSA), all comparisons were between all subjects, regardless of repeat / initial biopsy status

[c] Reported p-value for the comparison of the two AUCs computed among the same set of men

Table K3. Comparison of PCA3 and percent free PSA (%fPSA) differences in central estimates in men with positive and negative prostate biopsy results, after accounting for study-specific variability in measurements

Study/ Author[b]	Year	N	PCA3 Median for Pos Bx	PCA3 Median for Neg Bx	PCA3 Median Pooled Log SD	PCA3 Z_{PCA3}[a]	%fPSA Median for Pos Bx	%fPSA Median for Neg Bx	%fPSA Median Pooled Log SD	%fPSA $Z_{\%fPSA}$[a]	$Z_{PCA} - Z_{\%fPSA}$
Auprich[2]	2011	127	75	35	0.2929	1.00	11.0	17.0	0.1624	1.09	-0.91
Perdona[6]	2011	218	72	22	0.4264	1.21	15.0	17.0	0.1543	0.35	0.86
de la Taille[7]	2011	516	50	18	-	-	14.0	17.8			
Bollito[2]	2012	1246	63	35	0.4530	0.57	13.2	16.0	0.1858	0.42	0.15
All		**2107**									

Bx=prostate biopsy; Pos = positive, Neg = negative
[a] Z score = (log (Pos median) – log (Neg median)) / pooled log SD
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

Table K4. A comparison of PCA3 and %fPSA in identifying a positive prostate biopsy among matched studies: difference in sensitivities at the fixed specific associated with the commonly used PCA3 score cut-off of 35

Study / Author[b]	Year	Number	Initial Biopsy	Positive Biopsy	PCA3 score 1-Spec (%)	PCA3 score Sens A (%)	%fPSA[a] Sens B (%)	Diff (A-B) (%)
de la Taille[7]	2011	516	100%	40%	24.0	64.0	47.0	17

Spec=specificity, Sens = sensitivity; Diff = Difference = (Sens A – Sens B)
[a] Sensitivity for %fPSA elevation at the same false positive rate (1-specificity) found for a PCA3 score at a cut-off of 35.
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

PCA3 and %fPSA: Reported Medians and Standard Deviations

Four studies[2, 3, 6, 7] provided at least some relevant data that is presented in Table K3, sorted by study size. Two of the studies[6, 7] focus on the 'grey zone' of tPSA. One did not provide sufficient data to compute the z-scores[7], but did provide the median levels of both markers in those with positive versus negative biopsies. Among the three studies with a z-score difference, one found %fPSA to be slightly better[2], one found PCA3 to be slightly better[6] and one found only a small difference[3]. Due to the limited number of studies, no further analyses are presented.

PCA3 and %fPSA: Performance at a PCA3 Cut-Off Score of 35

Only one study[7] provided sufficient data to compare PCA3 sensitivity with the sensitivity of %fPSA at the specificity defined by the PCA3 cut-off score of 35 (Table K4). That study reported the PCA3 sensitivity to be 17 percent higher than the corresponding %fPSA sensitivity at the same (1-specificity) rate of 24 percent. With only a single study, no further analyses are presented.

PCA3 and %fPSA: ROC Curves - Sensitivity / Specificity

Five studies[1, 3-5, 7] provided sufficient data to compare PCA3 sensitivity versus %fPSA sensitivity at (1-specificity) rates of 20 percent through 80 percent (Table K5). For one study[3] two separate estimates were made; one for those men who were undergoing an initial biopsy, and another for those having a repeat biopsy. Although the majority of observations found PCA3 to have higher sensitivities at fixed specificities, half of six datasets found at least one point on the ROC curve at which %fPSA was better than PCA3. Due to the variability in the study populations and limited number of studies, no further analyses are presented.

Table K5. Sensitivity Differences (PCA3 - %fPSA) at PCA3 False positive rates (1 – Specificity)[a] from 20% to 80% in matched studies to identify positive biopsy men

Study/Author[b]	Year	Number	20%[c]	30%[c]	40%[c]	50%[c]	60%[c]	70%[c]	80%[c]
Ferro[4]	2012	151	45 (-5)	57 (-1)	79 (19)	87 (19)	91 (12)	96 (17)	100 (15)
Ploussard[5]	2010	301	40 (9)	56 (17)	71 (23)	80 (18)	83 (14)	89 (11)	93 (3)
Bollito[3] (repeat)	2012	509	27 (-3)	45 (0)	50 (-7)	57 (-9)	69 (-2)	76 (-1)	85 (-3)
de la Taille[7]	2011	516	57 (21)	71 (20)	77 (18)	85 (22)	89 (19)	93 (16)	95 (9)
Bollito[3] (initial)	2012	729	40 (9)	51 (7)	58 (6)	61 (1)	67 (-4)	79 (-2)	89 (3)
Aubin[1]	2010	1072	49 (13)	60 (13)	68 (10)	74 (3)	79 (0)	85 (0)	94 (1)
Median		**3286**	**43 (9)**	**57 (10)**	**70 (14)**	**77 (10)**	**81 (6)**	**87 (5)**	**94 (3)**

[a] (PCA3 sensitivity – %fPSA sensitivity) when (1-specificity) is held constant at values ranging from 20% to 80%.
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.
[c] Results presented as **PCA3 sensitivity (difference PCA3 sensitivity - %fPSA sensitivity)**. For example, in the first line (Ferro et al.), at a false positive rate (1 – specificity) of 20%, the PCA3 sensitivity is 45% and the sensitivity difference is -5% (e.g., PCA3 sensitivity 45% - %fPSA sensitivity 50% = -5%).

PCA3 and %fPSA: Regression Analysis

To be included in this analysis, studies would have reported the odds ratio (OR) for each marker, when that marker was assumed to be a continuous variable. The coefficients were used to compute the ratio of the ORs at the 25th and 75th centiles for each marker. This ratio of ORs for PCA3 will then be divided by the corresponding ratio for %fPSA. Values greater than 1 indicate PCA3 provides more discrimination than %fPSA. This normalization also allows for

comparisons between studies, where the coefficient is dependent on the range of %fPSA values studied.

Three studies[1, 6, 7] reported results of regression analysis separately for PCA3 and for %fPSA elevations in the same population of men. Only two of the studies[6, 7] provided sufficient data to complete the analysis (Table K6). Both found the ratio of ORs over the inter-quartile range to be greater for PCA3 than for %fPSA. All three studies included in this analysis restricted tPSA levels to < 10 ng/mL ('grey zone').

Table K6. Comparison of modeled univariate continuous odds ratios (OR) for PCA3 and %fPSA in matched studies

Study/ Author[a]	Year	N	PCA3 Report	OR @25th	OR @75th	Ratio A	%fPSA Report	OR @75th	OR @25th	Ratio B	Ratio (A/B)
Perdona[6]	2011	218	1.030	16	70	4.93	0.935	20	12	1.71	2.88
de la Taille[7]	2011	516	1.020	13	52	2.16	0.970	20[b]	12[b]	1.28	1.70
Aubin[1]	2010	1072	1.019	9	35	1.63	0.924	-	-	-	-
All		**1806**									

Bx = prostate biopsy, Corr = Correlation coefficient, OR = odds ratio, raw = as observed, adj = after adjustment for other marker
[a] Shaded rows indicate studies focusing on the 'grey zone' of tPSA; both excluded patients with tPSA levels of 10 ng/mL or greater.
[b] Centiles from Perdona[22] with no changes (medians identical).

Two specific measures were sought regarding the independence of PCA3 and %fPSA as markers for prostate biopsy status. Bivariate correlations (parametric or non-parametric) and results of logistic regression. Other markers may also be included in the regression model (*e.g.*, prostate volume). If both PCA3 and %fPSA coefficients remain essentially constant after adjusting for the other marker (and possibly, additional markers), this can be taken as evidence that the two markers together are more predictive that either alone.

Only two studies[1, 6] provided some information regarding independence of PCA3 and %fPSA (Table K7). None reported information on correlation coefficients. Both reported the univariate and multivariate ORs for PCA3 which indicated little change when other makers were added. Similar results were found for %fPSA, but in one of the studies[6], %fPSA was not included in the 'best' predictive model.

Table K7. Measures of independence of PCA3 and %fPSA in identifying men having a positive biopsy finding in matched studies

Study/ Author[a]	Year	N	Corr PCA3 & %fPSA in Pos Bx	Corr PCA3 & %fPSA in Neg Bx	PCA3 OR (raw, adj)	%fPSA OR (raw, adj)	Other relevant findings
Perdona[6]	2011	218	-	-	1.030/1.030	0.936/0.968	%fPSA not included in 'best' model
Aubin[1]	2010	1072	-	-	1.019/1.015	0.935/0.934	Model includes PCA3 values
All		**1290**					

Bx = prostate biopsy, Corr = Correlation coefficient, OR = odds ratio, raw = as observed, adj = after adjustment for other marker; Pos=positive; Neg=negative
[a] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

PCA3 and %fPSA Diagnostic Accuracy

PCA3 and %fPSA GRADE Strength of Evidence: Insufficient

- Risk of Bias: HIGH
  %fPSA is not highly correlated with tPSA and, therefore, the partial verification bias and sampling bias identified for the tPSA analyses may not have a strong influence on the %fPSA analyses. However, studies are observational and rated poor.
- Consistency: INCONSISTENT
  In general, 'grey zone' studies find PCA3 consistently higher while more generalizable studies find similar performance for the two markers (or find %fPSA slightly better). Given the small number of studies, it is not clear whether this is chance or that the %fPSA is subject to a higher bias than expected.
- Directness: INDIRECT
  The ultimate outcome of interest is long-term morbidity / mortality, and diagnostic accuracy is an intermediate outcome that cannot be linked directly to health outcomes.
- Precision: IMPRECISE
  A formal analysis of precision (e.g., confidence intervals) could not be perfored due to the matched nature of our analyses, the lack of original data and the limited number of included studies.

PCA3 and %fPSA Intermediate Outcome: Impact on Decisionmaking

No studies were identified that reported PCA3 and %fPSA levels along with outcomes related to the impact of decisionmaking on initial or repeat biopsy, or changing decisions leading to a reduction in unnecessary biopsies (false positives) with maintenance or improvement in cancer detection.

GRADE Strength of Evidence: Insufficient

PCA3 and %fPSA Intermediate Outcome: Biopsy-Related Harms

No studies were identified that reported PCA3 and %fPSA levels along with health outcomes related to harms of biopsy.

GRADE Strength of Evidence: Insufficient

PCA3 and %fPSA Long-Term Health Outcome: Morbidity/mortality

No studies were identified that reported PCA3 and %fPSA levels along with other long-term health outcomes, such as morbidity, mortality or quality of life.

GRADE Strength of Evidence: Insufficient

Comparator: Serum PSA Density (PSAD)

Intermediate Outcome: Diagnostic Accuracy

Key Points

PSAD measurements were compared with PCA3 scores to determine their diagnostic accuracy to predict prostate biopsy results (cancer/no cancer). Measures included in the analyses are the sensitivity, specificity (or the false positive rate equal to 1-specificity), and positive and negative predictive values.

- Area under the curve (AUC). Three studies[7, 9, 10] reported AUC estimates for PSAD and PCA3 in the same population. All found PCA3 scores to perform better than PSAD. However, two[7, 8] of the three focused on the 'grey zone' of tPSA results. Given the potential for the PSAD performance to be reduced when enrollees with higher tPSA results are excluded, findings must, at best, be restricted to the 'grey zone' setting.
- Reported medians and standard deviations (SD). Two studies[7, 8] provided data, but only one[8] provided sufficient data for analysis. The difference, reported as a z-score of 0.51, was in favor of PCA3.
- Performance at a PCA3 cut-off score of 35. Two studies[7, 10] reported the sensitivity and specificity of PCA3 at this cut-off and could be compared with PSAD sensitivity. One 'grey zone' study[7] found PCA3 to be 17 percent higher, while one study in the general population[9] found PCA3 to be 11 percent lower.
- ROC Curves - Sensitivity / Specificity. Three studies[7-9] provided a ROC curve, or data representing a ROC curve for both markers. Two[7, 8] were 'grey zone' studies. At a specificity of 50 percent, the differences in corresponding (PCA3 – PSAD) sensitivities ranged from 3 to 12 percent.
- Regression analysis. Only one study[10] provided data regarding the correlation between PCA3 and PSAD among biopsy positive men, and the estimate was low (r = 0.13).

Study Characteristics

Four studies[7-10] reported PCA3 and PSAD comparisons that could be used in one or more of the matched analyses (Table K8). Overall, the analyses were inconsistent in finding whether the PCA3 or PSAD measurements were better at identifying positive prostate biopsies, and any findings should be restricted to the 'grey zone' setting.

Table K8. Summary results for the five available analytic comparisons[a] of PCA3 versus PSAD in matched populations of men having prostate biopsies

Study/ Author[b]	Year	N	AUC	Mean/SD	PCA3>35	Sens/Spec	Reg
Wu[9]	2012	103	0.0640	-	-10%	3%	-
de la Taille[7]	2011	516	0.0720	-	17%	12%	-
Durand[10]	2012	160	-	-	-	-	r=0.13
Ochiai[8]	2011	105	0.0342	0.51	-	9%	-
All		**884**					

[a] AUC = area under the curve for PCA3 minus the AUC of tPSA; Mean/DS = difference in separation between PCA3 scores and tPSA results, when expressed as z-scores; PCA3>35 = difference of the PCA3 minus the tPSA sensitivities at the specificity found for a PCA3 cut-off of 35; Sens/Spec = difference between PCA3 and tPSA sensitivity at a specificity of 50%; Reg = relative change in the PCA3 ORs (between the 25th and 75th centiles) and the corresponding tPSA ORs. The corresponding full analyses resulting in these summaries can be found on the following pages.

[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

PCA3 and PSAD: Area Under the Curve

Three studies[7-9] satisfied inclusion criteria and relevant data are presented in Table K9. Two found PCA3 to have a higher AUC, but both were 'grey zone' studies.[7, 8] The third study[9] allowed participants with high tPSA to be enrolled, and it found the two AUCs to be similar (-0.04). Due to the differences in the study populations and the limited number of studies, no further analyses are presented.

Table K9. Comparing PCA3 levels and PSA density (PSAD) measurements in matched studies via AUC analysis to correctly diagnose prostate cancer, as defined by a positive biopsy

Author[a]	Year	Number	Initial Bx	PCA3 AUC	PSAD AUC	Difference[b]	P-value[c]
Wu[9]	2012	103	0%	0.6400	0.6800	-0.0400	-
Ochiai[8]	2011	105	81%	0.8507	0.8164	0.0342	0.67
de la Taille[7]	2011	516	100%	0.7610	0.6890	0.0720	0.023
All		**724**					

Bx = biopsy, AUC = area under the curve, '-' = no value reported
[a] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.
[b] PCA3 AUC – PSAD AUC
[c] Reported p-value for the comparison of the two AUCs computed among the same set of men

PCA3 and PSAD: Reported Medians and Standard Deviations

Two studies[7, 8] satisfied the inclusion criteria, and relevant data are presented in Table K10. In one study[8], the log standard deviation was estimated using the range of results divided by six. The difference in z-scores was 0.51, indicating a better separation between men with positive and negative biopsies for PCA3 compared to PSAD. In the other study[7], it was not possible to estimate the log standard deviation. Due to the single study, no further analyses are presented.

PCA3 and PSAD: Performance at a PCA3 Cut-Off Score of 35

Two studies[7, 9] provided sufficient data to compare PCA3 sensitivity with the sensitivity of PSAD at the specificity defined by the PCA3 cut-off score of 35 (Table K11). One study[8] found PCA3 to be better, but that study focused on the 'grey zone'. The second study[9] was in the general population of tPSA/DRE positive patients, and found PSAD to be better. Due to the limited number of studies, no further analyses are presented.

PCA3 and PSAD: ROC Curves - Sensitivity / Specificity

Three studies[7-9] provide sufficient data to compare PCA3 sensitivity versus PSAD sensitivity at (1-specificity) rates of 20 percent through 80 percent (Table K12). The results show the two tests to be similar, with point estimates sometimes higher for PCA3 and sometimes for PSAD. However, due to the differences in the populations studied and the limited number of studies, no further analyses are presented.

PCA3 and PSAD: Regression Analysis

Only one study[10] reported results of the selected analyses regarding the independence of PCA3 and PSAD in predicting prostate biopsy results, and it only included cases (Table K14). The correlation was low (0.13).

Table K10. Comparison of PCA3 and PSA density (PSAD) in central estimates in men with positive and negative prostate biopsy results, after accounting for study-specific variability in measurements

Study/ Author[b]	Year	N	PCA3 Median for Pos Bx	PCA3 Median for Neg Bx	PCA3 Median Pooled Log SD	PCA3 Z_{PCA3}[a]	PSAD (ng/mL) Median for Pos Bx	PSAD (ng/mL) Median for Neg Bx	PSAD Median Pooled Log SD	PSAD Z_{PSAD}[a]	$Z_{PCA} - Z_{PSAD}$
Ochiai[8]	2011	105	59.5	14.2	0.3489	1.78	0.38	0.17	0.2740	1.27	0.51
de la Taille[7]	2011	516	50.0	18.0	-	-	0.15	0.11	-	-	-
All		**621**									

Bx = prostate biopsy, Pos = positive, Neg = negative
[a] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.
[b] Z score = (log (Pos median) – log (Neg median)) / pooled log SD

Table K11. A comparison of PCA3 and PSAD in identifying a positive prostate biopsy among matched studies: difference in Sensitivities at the fixed specific associated with the commonly used PCA3 score cut-off of 35

Study/Author[b]	Year	Number	Initial Bx	Positive Bx	PCA3 Score 1-Spec (%)	PCA3 Sens (%) A	PSAD[a] Sens (%) B	Difference (A-B) (%)
Wu[9]	2012	103	0%	36.0%	23.0	38	49	-11%
de la Taille[7]	2011	516	100%	40.0%	24.0	64	47	17%
All		**619**						

Bx=prostate biopsy; Spec=specificity (1-specificity=false positive rate), Sens = sensitivity (detection rate)
[a] Sensitivity for PSAD elevation at the same false positive rate (1-specificity) found for a PCA3 score at a cut-off of 35.
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

Table K12. Sensitivity Differences (PCA3 - PSAD) at PCA3 False positive rates (1 – Specificity)[a] from 20% to 80% in matched studies to identify positive biopsy men

Study/ Author[b]	Year	Number	20%[c]	30%[c]	40%[c]	50%[c]	60%[c]	70%[c]	80%[c]
Wu[9]	2012	103	37 (- 8)	45 (-14)	67 (- 1)	77 (3)	82 (4)	82 (- 3)	82 (- 3)
de la Taille[7]	2011	516	57 (13)	71 (15)	77 (15)	85 (12)	89 (3)	93 (3)	95 (0)
Ochiai[8]	2011	105	74 (- 3)	87 (3)	92 (3)	97 (9)	98 (4)	100 (0)	100 (0)
Median		**724**	**57 (- 3)**	**71 (3)**	**77 (3)**	**85 (9)**	**89 (4)**	**93 (0)**	**95 (0)**

[a] (PCA3 sensitivity – PSAD sensitivity) when (1-specificity) is held constant at values ranging from 20% to 80%.
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.
[c] Results presented as **PCA3 sensitivity (difference PCA3 sensitivity - EVN sensitivity)**.

Table K13. Summary results for the five analytic comparisons[a] of PCA3 versus PSAD in matched populations of men having prostate biopsies

Study/Author[b]	Year	N	AUC	Mean/SD	PCA3>35	Sens/Spec	Reg
Ochiai[8]	2011	105	-	-	[c]	-	-
Perdona[6]	2011	218	0.1130	-	0.52, 0.22[d]	88 (4)	Increase in AUC
Ankerst[12]	2008	443	0.0120	-	-	-	-
FDA Summary[11]	2012	464	0.0540	-	-	-	Increase in AUC
All		**1230**					

[a] AUC = area under the curve for PCA3 minus the AUC of tPSA; Mean/DS = difference in separation between PCA3 scores and tPSA results, when expressed as z-scores; PCA3>35 = difference of the PCA3 minus the tPSA sensitivities at the specificity found for a PCA3 cut-off of 35; Sens/Spec = difference between PCA3 and tPSA sensitivity at a specificity of 50%; Reg = relative change in the PCA3 ORs (between the 25th and 75th centiles) and the corresponding tPSA ORs. The corresponding full analyses resulting in these summaries can be found on the following pages.
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.
[c] Did not provide ROC information above a false positive rate of 40%, but the EVN was consistently higher.
[d] The two entries are for the Chun risk model and the PCPT model, respectively.

PCA3 and PSAD Intermediate Outcome: Diagnostic Accuracy

- Risk of Bias: HIGH.
  PSAD does not appear to be highly correlated with tPSA and, therefore, the partial verification bias, and sampling bias identified for the tPSA analyses are not expected to have a strong effect on the PSAD analyses. However, studies were observational and rated poor.
- Consistency: UNKNOWN
  Two few data were available to assess consistency.
- Directness: INDIRECT
  The ultimate outcome of interest is long-term morbidity / mortality, and diagnostic accuracy is an intermediate outcome that cannot be linked directly to health outcomes.
- Precision: IMPRECISE
  A formal analysis of precision (e.g., confidence intervals) was not able to be computed due to the matched nature of our analyses, the lack of original data, and the limited number of included studies.

GRADE Strength of Evidence: Insufficient

PCA3 and PSAD - Intermediate Outcome: Impact on Decisionmaking

No studies were identified that reported PCA3 and PSAD levels along with outcomes related to the impact of decisionmaking on initial or repeat biopsy, or changing decisions leading to a reduction in unnecessary biopsies (false positives) with maintenance or improvement in cancer detection.

GRADE Strength of Evidence: Insufficient

PCA3 and PSAD - Intermediate Outcome: Biopsy-Related Harms

No studies were identified that reported PCA3 and PSAD levels along with health outcomes related to harms of biopsy.

GRADE Strength of Evidence: Insufficient

PCA3 and PSAD - Long-Term Health Outcome: Morbidity/Mortality

No studies were identified that reported PCA3 and PSAD levels along with long-term health outcomes, such as morbidity, mortality or quality of life.

GRADE Strength of Evidence: Insufficient

Comparator: Externally Validated Nomogram (EVN)

PCA3 and EVN Intermediate Outcome: Diagnostic Accuracy

Key points

Externally validated nomograms were compared with PCA3 scores to determine their diagnostic accuracy to predict prostate biopsy results (cancer / no cancer). Measures included in the analyses are the sensitivity, specificity (or the false positive rate equal to 1-specificity), and positive and negative predictive values.

- Area under the curve (AUC). Three studies[6, 11, 12] reported AUC estimates for EVN and PCA3 in the same population (Table K14). Both found the better performance for PCA3 (AUC differences of 0.01 and 0.11).
- Reported medians and standard deviations (SD). No studies provided sufficient data for analysis
- Performance at a PCA3 cut-off score of 35. One study[6] reported comparative results for PCA3 versus two separate EVN algorithms.
- ROC Curves - Sensitivity / Specificity. Two studies[6, 8] provided a ROC curve, or data representing a ROC curve for both markers. Only one of these[6] reported data over the required range of specificities (20% to 80%). The second provided only limited data, with no information at the specificity of 50 percent.
- Regression analysis. One study[11] reported results and did not find a significant association.

Study Characteristics

Four studies[6, 8, 11, 12] reported PCA3 and EVN comparisons that could be used in one or more of the matched analyses (Table K13). Comparisons included PCA3 to Chun's nomogram, the Prostate Cancer Prevention Trial (PCPT) model and other multivariate analyses that included variables such as tPSA level, age, DRE result, family history or race. Of the seven comparisons,

all found PCA3 performed better than PSAD. The following sections provide detailed information about the five analyses performed and the specific findings.

PCA3 and EVN: Area Under the Curve

Three studies[6, 12 11] satisfied inclusion criteria and relevant data are presented in Table K14. Both found that the AUC of PCA3 was greater than for EVN, but the smaller difference found for one study[12] did not reach statistical significance. Due to the limited number of studies, no further analyses are presented.

Table K14. Comparing PCA3 and EVN scores in matched studies via AUC analysis to correctly diagnose prostate cancer, as defined by a positive needle biopsy

Author[a]	Year	Number	Initial Bx	PCA3 AUC	EVN AUC	Difference[b]	P-value[c]
Perdona[6]	2011	218	61%	0.8280	0.7150	0.1130	0.003
Ankerst[12]	2008	443	81%	0.6650	0.6530	0.0120	NS
FDA Summary[11]	2012	464	0%	0.7070	0.6530	0.0540	<0.05
All		**1125**					

Bx = biopsy, AUC = area under the curve, '-' = no value reported
[a] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.
[b] PCA3 AUC – EVN AUC
[c] Reported p-value for the comparison of the two AUCs computed among the same set of men

PCA3 and EVN: Reported Medians and Standard Deviations

No studies were identified that provided relevant data.

PCA3 and EVN: Performance at a PCA3 Cut-Off Score of 35

One study[6] satisfied the inclusion criteria and relevant data are presented in Table K15. That group used the Chun's and PCPT risk nomograms. Both found that PCA3 was better than either of the externally validated nomograms. Due to the limited number of studies, no further analyses are presented.

PCA3 and EVN: ROC Curves - Sensitivity / Specificity

Two studies[6, 8] provide sufficient data to compare PCA3 sensitivity versus EVN sensitivity at (1-specificity) rates. Both used the risk algorithm by Chun. One[6] provided data for specificities of 20 percent through 80 percent (Table K16) and found PCA3 to be better. The second study[8] only reported results for 20 percent through 40 percent and found the EVN only slightly better. Due to the limited number of studies, no further analyses are presented.

PCA3 and EVN: Regression analysis

Two studies provided a comparison of an externally validated model with and without PCA3.[11,12] In the first study[12], the PCPT AUC increased from 0.653 to 0.696 with the inclusion of PCA3. However, this increase was not statistically significant. In the second study[11], the inclusion of a dichotomized PCA3 score to tPSA and 'standard of care covariates' (age, DRE result, family history, race, number previous negative biopsies) resulted in an increased AUC from 0.707 to 0.733.

PCA3 and EVN - Intermediate Outcome: Diagnostic Accuracy

- Risk of Bias: HIGH.
  EVNs often include tPSA and, therefore, the partial verification bias, and sampling bias identified for the tPSA analyses may have an impact. However, none of the models actually used tPSA (which are most subject to bias) and thus biasing the EVN analyses is not expected. However, studies were observational and rated poor.
- Consistency: UNKNOWN
  Two few data were available to assess consistency.
- Directness: INDIRECT
  The ultimate outcome of interest is long-term morbidity / mortality, and diagnostic accuracy is an intermediate outcome that cannot be linked directly to health outcomes.
- Precision: IMPRECISE
  A formal analysis of precision (e.g., confidence intervals) was not able to be computed due to the matched nature of our analyses, the lack of original data, and the limited number of included studies.

GRADE Strength of Evidence: Insufficient

PCA3 and EVN - Intermediate Outcome: Impact on Decisionmaking

No studies were identified that reported PCA3 and EVN results along with outcomes related to the impact of decisionmaking on initial or repeat biopsy, or changing decisions leading to a reduction in unnecessary biopsies (false positives) with maintenance or improvement in cancer detection.

GRADE Strength of Evidence: Insufficient

PCA3 and EVN - Intermediate Outcome: Biopsy-Related Harms

No studies were identified that reported PCA3 and EVN results along with health outcomes related to harms of biopsy.

GRADE Strength of Evidence: Insufficient

PCA3 and EVN - Long-Term Health Outcome: Morbidity/Mortality

No studies were identified that reported PCA3 and EVN results along with long-term health outcomes, such as morbidity, mortality or quality of life.

GRADE Strength of Evidence: Insufficient

Table K15. Comparison of PCA3 and externally validated nomograms (EVN) in central estimates in men with positive and negative prostate biopsy results, after accounting for study-specific variability in measurements

Study/ Author[b]	Year	N	EVN	PCA3 score Median for Pos Bx	PCA3 score Median for Neg Bx	PCA3 score Median Pooled Log SD	PCA3 Z_{PCA3}[a]	EVN score Median for Pos Bx	EVN score Median for Neg Bx	EVN score Median Pooled Log SD	EVN Z_{EVN}[a]	$Z_{PCA} - Z_{EVN}$
Perdona[6]	2011	218	Chun's "risk"	72.0	22.0	0.4264	1.21	54	41	18.9	0.69	0.52
Perdona[6]	2011	218	PCPT "risk"	72.0	22.0	0.4264	1.21	54	39	15.2	0.99	0.22

Bx = prostate biopsy. Pos = prostate biopsy positive, Neg = prostate biopsy negative; Bx = prostate biopsy, SD = standard deviation
[a] Z score = (log (Pos median) – log (Neg median)) / pooled log SD
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

Table K16. Sensitivity Differences (PCA3 - EVN) at PCA3 False positive rates (1 – Specificity)[a] from 20% to 80% in matched studies to identify positive biopsy men

Study/Author[b]	Year	N	20%[c]	30%[c]	40%[c]	50%[c]	60%[c]	70%[c]	80%[c]
Perdona[6]	2011	218	70 (23)	79 (27)	82 (20)	88 (4)	97 (4)	100 (4)	100 (2)
Ochiai[8]	2011	105	44 (- 7)	56 (- 6)	67 (- 4)	-	-	-	-

[a] (PCA3 sensitivity – EVN sensitivity) when (1-specificity) is held constant at values ranging from 20% to 80%.
[b] Shaded rows indicate studies focusing on the 'grey zone' of tPSA.
[c] Results presented as **PCA3 sensitivity (difference PCA3 sensitivity - EVN sensitivity)**.

Comparator: Complexed PSA (cPSA)

PCA3 and cPSA - Intermediate Outcome: Diagnostic Accuracy

No studies were identified that provide matched data for PCA3 and cPSA levels in eligible study populations with biopsy results.

GRADE Strength of Evidence: Insufficient

PCA3 and cPSA - Intermediate Outcome: Impact on Decisionmaking

No studies were identified that reported PCA3 and cPSA levels along with outcomes related to the impact of decisionmaking on initial or repeat biopsy, or changing decisions leading to a reduction in unnecessary biopsies (false positives) with maintenance or improvement in cancer detection.

GRADE Strength of Evidence: Insufficient

PCA3 and cPSA - Intermediate Outcome: Biopsy-Related Harms

No studies were identified that reported PCA3 and cPSA levels along with health outcomes related to harms of biopsy.

GRADE Strength of Evidence: Insufficient

PCA3 and cPSA - Long-term Health Outcome: Morbidity/Mortality

No studies were identified that reported PCA3 and cPSA levels along with long-term health outcomes, such as morbidity, mortality or quality of life.

GRADE Strength of Evidence: Insufficient

Comparator: Total PSA Doubling Time (DT) and Total PSA Velocity (PSAV)

PCA3 and DT/PSAV - Intermediate Outcome: Diagnostic Accuracy

No studies were identified that provide matched data for PCA3 and DT/PSAV results levels in eligible study populations with biopsy results.

GRADE Strength of Evidence: Insufficient

PCA3 and DT/PSAV - Intermediate Outcome: Impact on Decisionmaking

No studies were identified that reported PCA3 and DT/PSAV results along with outcomes related to the impact of decisionmaking on initial or repeat biopsy, or changing decisions leading to a reduction in unnecessary biopsies (false positives) with maintenance or improvement in cancer detection.

GRADE Strength of Evidence: Insufficient

PCA3 and DT/PSAV - Intermediate Outcome: Biopsy-Related Harms

No studies were identified that reported PCA3 and DT/PSAV results along with health outcomes related to harms of biopsy.

GRADE Strength of Evidence: Insufficient

PCA3 and DT/PSAV - Long-term Health Outcome: Morbidity/Mortality

Key points

No studies were identified that reported PCA3 and DT/PSAV results along with long-term health outcomes, such as morbidity, mortality or quality of life.

GRADE Strength of Evidence: Insufficient

Comparator: Multivariate Models including PCA3 and tPSA

Study Characteristics

Three studies[1, 6, 7] reported multivariate models (logistic regression) that included both tPSA elevations and PCA3 scores. Only one of these[7] provided results for both a base model and the base model plus PCA3 scores. The base model included the man's age, DRE (categorical), tPSA (continuous), and prostate volume (continuous); all were statistically significant ($p<0.01$). PCA3 score as a continuous variable was then included. The coefficients in the base model were essentially unchanged with all point estimates being within the 95 percent CI of the base model estimates and all remained statistically significant. The p-value associated with PCA3 measurements was <0.001. This implied that the PCA3 scores added independent information.

However, there are several limitations to the study. Neither the tPSA nor PCA3 measurements were reported to have been transformed. An underlying assumption of logistic regression is that continuous variables are reasonably Gaussian and this was most likely violated by the modeling. Second, the authors provided no information regarding the change in sensitivity at fixed specificities. Rather, they provided a modest increase in predictive accuracy, a difficult estimate to interpret.

Multivariate Models Including PCA3 and tPSA - Outcome: Diagnostic Accuracy

GRADE Strength of Evidence: Insufficient

- Risk of Bias: HIGH.
  Multivariate models often include tPSA and, therefore, the partial verification bias, and sampling bias identified for the tPSA analyses may have a modest impact. However, studies were observational and rated poor.
- Consistency: UNKNOWN
  The limited number of studies (three) did not report results in a consistent manner (e.g., correlations or effect sizes).
- Directness: INDIRECT
  The ultimate outcome of interest is long-term morbidity / mortality; the subject of Key Question 3.
- Precision: IMPRECISE
  A formal analysis of precision (e.g., confidence intervals) was not able to be computed due to the matched nature of our analyses, the lack of original data, and the limited number of included studies.

Multivariate Models Including PCA3 and tPSA - Intermediate Outcome: Impact on Decisionmaking

No studies were identified that reported the multivariate model results along with outcomes related to the impact of decisionmaking on initial or repeat biopsy, or changing decisions leading to a reduction in unnecessary biopsies (false positives) with maintenance or improvement in cancer detection.

GRADE Strength of Evidence: Insufficient

Multivariate Models Including PCA3 and tPSA - Intermediate Outcome: Biopsy-Related Harms

No studies were identified that reported the multivariate model results along with health outcomes related to harms of biopsy.

GRADE Strength of Evidence: Insufficient

Multivariate Models Including PCA3 and tPSA - Long-term Health Outcome: Morbidity/Mortality

No studies were identified that reported the multivariate model results along with long-term health outcomes, such as morbidity, mortality or quality of life.

GRADE Strength of Evidence: Insufficient

References

1. Aubin SM, Reid J, Sarno MJ, et al. PCA3 molecular urine test for predicting repeat prostate biopsy outcome in populations at risk: validation in the placebo arm of the dutasteride REDUCE trial. J Urol. 2010 Nov;184(5):1947-52. PMID: 20850153.
2. Auprich M, Augustin H, Budaus L, et al. A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. BJU Int. 2012 Jun;109(11):1627-35. PMID: 21939492.
3. Bollito E, De Luca S, Cicilano M, et al. Prostate cancer gene 3 urine assay cutoff in diagnosis of prostate cancer: a validation study on an Italian patient population undergoing first and repeat biopsy. Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology. 2012 Apr;34(2):96-104. PMID: 22611765.
4. Ferro M, Bruzzese D, Perdona S, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta. 2012 Aug 16;413(15-16):1274-8. PMID: 22542564.
5. Ploussard G, Haese A, Van Poppel H, et al. The prostate cancer gene 3 (PCA3) urine test in men with previous negative biopsies: does free-to-total prostate-specific antigen ratio influence the performance of the PCA3 score in predicting positive biopsies? BJU Int. 2010 Oct;106(8):1143-7. PMID: 20230386.
6. Perdona S, Cavadas V, Di Lorenzo G, et al. Prostate cancer detection in the "grey area" of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. European urology. 2011 Jan;59(1):81-7. PMID: 20947244.
7. de la Taille A, Irani J, Graefen M, et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. J Urol. 2011 Jun;185(6):2119-25. PMID: 21496856.
8. Ochiai A, Okihara K, Kamoi K, et al. Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. Int J Urol. 2011 Mar;18(3):200-5. PMID: 21332814.
9. Wu AK, Reese AC, Cooperberg MR, et al. Utility of PCA3 in patients undergoing repeat biopsy for prostate cancer. Prostate Cancer Prostatic Dis. 2012 Mar;15(1):100-5. PMID: 22042252.
10. Durand X, Xylinas E, Radulescu C, et al. The value of urinary prostate cancer gene 3 (PCA3) scores in predicting pathological features at radical prostatectomy. BJU Int. 2012 Jul;110(1):43-49. PMID: 22221521.
11. U.S. Food and Drug Administration. Summary of Safety and Effectiveness Data: PROGENSA PCA3 Assay, 2012. Accessed July 16, 2012. http://www.accessdata.fda.gov/cdrh_docs/pdf10/P100033b.pdf.
12. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008 Oct;180(4):1303-8; discussion 08. PMID: 18707724.